Activity Report Saclay - Île-de-France 2019

Edition: 2020-03-23

List of Inria's Research Teams

Project-Team AVIZ

Analysis and Visualization

IN COLLABORATION WITH: Laboratoire de recherche en informatique (LRI)

IN PARTNERSHIP WITH:
CNRS

RESEARCH CENTER
Saclay - Île-de-France

THEME
Interaction and visualization

Table of contents

Project-Team AVIZ

Creation of the Team: 2007 February 08, updated into Project-Team: 2008 January 01

Keywords:

Computer Science and Digital Science:

A1.3. - Distributed Systems
A1.3.3. - Blockchain
A3.1.4. - Uncertain data
A3.1.7. - Open data
A3.1.8. - Big data (production, storage, transfer)
A3.3. - Data and knowledge analysis
A3.3.1. - On-line analytical processing
A3.3.3. - Big data analysis
A3.5.1. - Analysis of large graphs
A5.1. - Human-Computer Interaction
A5.1.1. - Engineering of interactive systems
A5.1.2. - Evaluation of interactive systems
A5.1.6. - Tangible interfaces
A5.1.8. - 3D User Interfaces
A5.1.9. - User and perceptual studies
A5.2. - Data visualization
A5.6.1. - Virtual reality
A5.6.2. - Augmented reality
A6.3.3. - Data processing
A9.6. - Decision support

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1. - Biology
B1.2. - Neuroscience and cognitive science
B9.5.6. - Data science
B9.6. - Humanities
B9.6.1. - Psychology
B9.6.3. - Economy, Finance
B9.6.6. - Archeology, History
B9.6.10. - Digital humanities

1. Team, Visitors, External Collaborators

Research Scientists

Jean-Daniel Fekete [Team leader, Inria, Senior Researcher, HDR]
Tobias Isenberg [Team leader, Inria, Senior Researcher, HDR]
Pierre Dragicevic [Inria, Researcher]

Steve Haroz [Inria, Starting Research Position]
Petra Isenberg [Inria, Researcher]
Catherine Plaisant [Inria, International Chair, Advanced Research Position]

External Collaborators

Paolo Buono [University of Bari, from Oct 2019]
Evelyne Lutton [INRA]
Frédéric Vernier [Univ Paris-Sud]
Christoph Kinkeldey [Freie Universität Berlin]

Technical Staff

Christian Poli [Inria, Engineer]

PhD Students

Yuheng Feng [Univ Paris-Sud, PhD Student, from Oct 2019]
Sarkis Halladjian [Inria, PhD Student]
Mohammad Alaul Islam [Inria, PhD Student, from Oct 2019]
Alexis Pister [Inria, PhD Student, from Oct 2019]
Mickael Sereno [Inria, PhD Student]
Natkamon Tovanich [Institut de recherche technologique System X, PhD Student]
Xiyao Wang [Inria, PhD Student]

Post-Doctoral Fellows

Tanja Blascheck [Inria, Post-Doctoral Fellow, until Feb 2019]
Qing Chen [École polytechnique, Post-Doctoral Fellow]
Paola Tatiana Llerena Valdivia [Inria, Post-Doctoral Fellow]
Gaëlle Richer [Inria, Post-Doctoral Fellow, from Dec 2019]

Visiting Scientists

Remi Bardes [Univ Paris-Sud, until Apr 2019]
Paolo Buono [University of Bari, from Aug 2019 until Sep 2019]
Claudio Silva [New York University, until Jul 2019]

Administrative Assistant

Katia Evrat [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Objectives

Aviz (Analysis and VIsualiZation) is a multidisciplinary project that seeks to improve visual exploration and analysis of large, complex datasets by tightly integrating analysis methods with interactive visualization.

Our work has the potential to affect practically all human activities for and during which data is collected and managed and subsequently needs to be understood. Often data-related activities are characterized by access to new data for which we have little or no prior knowledge of its inner structure and content. In these cases, we need to interactively *explore* the data first to gain insights and eventually be able to act upon the data contents. Interactive visual analysis is particularly useful in these cases where automatic analysis approaches fail and human capabilities need to be exploited and augmented.

Within this research scope Aviz focuses on five research themes:

- Methods to visualize and smoothly navigate through large datasets;
- Efficient analysis methods to reduce huge datasets to visualizable size;
- Visualization interaction using novel capabilities and modalities;
- Evaluation methods to assess the effectiveness of visualization and analysis methods and their usability;
- Engineering tools for building visual analytics systems that can access, search, visualize and analyze large datasets with smooth, interactive response.

2.2. Research Themes

Aviz's research on Visual Analytics is organized around five main Research Themes:

Methods to visualize and smoothly navigate through large data sets: Large data sets challenge current visualization and analysis methods. Understanding the structure of a graph with one million vertices is not just a matter of displaying the vertices on a screen and connecting them with lines. Current screens only have around two million pixels. Understanding a large graph requires both data reduction to visualize the whole and navigation techniques coupled with suitable representations to see the details. These representations, aggregation functions, navigation and interaction techniques must be chosen as a coordinated whole to be effective and fit the user's mental map.

Aviz designs new visualization representations and interactions to efficiently navigate and manipulate large data sets.

Efficient analysis methods to reduce huge data sets to visualizable size: Designing analysis components with interaction in mind has strong implications for both the algorithms and the processes they use. Some data reduction algorithms are suited to the principle of sampling, then extrapolating, assessing the quality and incrementally enhancing the computation: for example, all the linear reductions such as PCA, Factorial Analysis, and SVM, as well as general MDS and Self Organizing Maps. Aviz investigates the possible analysis processes according to the analyzed data types.

Visualization interaction using novel capabilities and modalities: The importance of interaction to Visualization and, in particular, to the interplay between interactivity and cognition is widely recognized. However, information visualization interactions have yet to take full advantage of these new possibilities in interaction technologies, as they largely still employ the traditional desktop, mouse, and keyboard setup of WIMP (Windows, Icons, Menus, and a Pointer) interfaces. At Aviz we investigate in particular interaction through tangible and touch-based interfaces to data.

Evaluation methods to assess their effectiveness and usability: For several reasons appropriate evaluation of visual analytics solutions is not trivial. First, visual analytics tools are often designed to be applicable to a variety of disciplines, for various different data sources, and data characteristics, and because of this variety it is hard to make general statements. Second, in visual analytics the specificity of humans, their work environment, and the data analysis tasks, form a multi-faceted evaluation context which is difficult to control and generalize. This means that recommendations for visual analytics solutions are never absolute, but depend on their context.

In our work we systematically connect evaluation approaches to visual analytics research—we strive to develop and use both novel as well as establish mixed-methods evaluation approaches to derive recommendations on the use of visual analytics tools and techniques. Aviz regularly published user studies of visual analytics and interaction techniques and takes part in dedicated workshops on evaluation.

Engineering tools: for building visual analytics systems that can access, search, visualize and analyze large data sets with smooth, interactive response.

Currently, databases, data analysis and visualization all use the concept of data tables made of tuples and linked by relations. However, databases are storage-oriented and do not describe the data types precisely. Analytical systems describe the data types precisely, but their data storage and computation model are not suited to interactive visualization. Visualization systems use in-memory data tables tailored for fast display and filtering, but their interactions with external analysis programs and databases are often slow.

Aviz seeks to merge three fields: databases, data analysis and visualization. Part of this merging involves using common abstractions and interoperable components. This is a long-term challenge, but it is a necessity because generic, loosely-coupled combinations will not achieve interactive performance. First progress has already been made [10].

Aviz's approach is holistic: these five themes are facets of building an analysis process optimized for discovery. All the systems and techniques Aviz designs support the process of understanding data and forming insights while minimizing disruptions during navigation and interaction.

3. Research Program

3.1. Scientific Foundations

The scientific foundations of Visual Analytics lie primarily in the domains of Visualization and Data Mining. Indirectly, it inherits from other established domains such as graphic design, Exploratory Data Analysis (EDA), statistics, Artificial Intelligence (AI), Human-Computer Interaction (HCI), and Psychology.

The use of graphic representation to understand abstract data is a goal Visual Analytics shares with Tukey's Exploratory Data Analysis (EDA) [67], graphic designers such as Bertin [54] and Tufte [66], and HCI researchers in the field of Information Visualization [53].

EDA is complementary to classical statistical analysis. Classical statistics starts from a *problem*, gathers *data*, designs a *model* and performs an *analysis* to reach a *conclusion* about whether the data follows the model. While EDA also starts with a problem and data, it is most useful *before* we have a model; rather, we perform visual analysis to discover what kind of model might apply to it. However, statistical validation is not always required with EDA; since often the results of visual analysis are sufficiently clear-cut that statistics are unnecessary.

Visual Analytics relies on a process similar to EDA, but expands its scope to include more sophisticated graphics and areas where considerable automated analysis is required before the visual analysis takes place. This richer data analysis has its roots in the domain of Data Mining, while the advanced graphics and interactive exploration techniques come from the scientific fields of Data Visualization and HCI, as well as the expertise of professions such as cartography and graphic designers who have long worked to create effective methods for graphically conveying information.

The books of the cartographer Bertin and the graphic designer Tufte are full of rules drawn from their experience about how the meaning of data can be best conveyed visually. Their purpose is to find effective visual representation that describe a data set but also (mainly for Bertin) to discover structure in the data by using the right mappings from abstract dimensions in the data to visual ones.

For the last 25 years, the field of Human-Computer Interaction (HCI) has also shown that interacting with visual representations of data in a tight perception-action loop improves the time and level of understanding of data sets. Information Visualization is the branch of HCI that has studied visual representations suitable to understanding and interaction methods suitable to navigating and drilling down on data. The scientific foundations of Information Visualization come from theories about perception, action and interaction.

Several theories of perception are related to information visualization such as the "Gestalt" principles, Gibson's theory of visual perception [59] and Triesman's "preattentive processing" theory [65]. We use them extensively but they only have a limited accuracy for predicting the effectiveness of novel visual representations in interactive settings.

Information Visualization emerged from HCI when researchers realized that interaction greatly enhanced the perception of visual representations.

To be effective, interaction should take place in an interactive loop faster than 100ms. For small data sets, it is not difficult to guarantee that analysis, visualization and interaction steps occur in this time, permitting smooth data analysis and navigation. For larger data sets, more computation should be performed to reduce the data size to a size that may be visualized effectively.

In 2002, we showed that the practical limit of InfoVis was on the order of 1 million items displayed on a screen [57]. Although screen technologies have improved rapidly since then, eventually we will be limited by the physiology of our vision system: about 20 millions receptor cells (rods and cones) on the retina. Another problem will be the limits of human visual attention, as suggested by our 2006 study on change blindness in large and multiple displays [55]. Therefore, visualization alone cannot let us understand very large data sets. Other techniques such as aggregation or sampling must be used to reduce the visual complexity of the data to the scale of human perception.

Abstracting data to reduce its size to what humans can understand is the goal of Data Mining research. It uses data analysis and machine learning techniques. The scientific foundations of these techniques revolve around the idea of finding a good model for the data. Unfortunately, the more sophisticated techniques for finding models are complex, and the algorithms can take a long time to run, making them unsuitable for an interactive environment. Furthermore, some models are too complex for humans to understand; so the results of data mining can be difficult or impossible to understand directly.

Unlike pure Data Mining systems, a Visual Analytics system provides analysis algorithms and processes compatible with human perception and understandable to human cognition. The analysis should provide understandable results quickly, even if they are not ideal. Instead of running to a predefined threshold, algorithms and programs should be designed to allow trading speed for quality and show the tradeoffs interactively. This is not a temporary requirement: it will be with us even when computers are much faster, because good quality algorithms are at least quadratic in time (e.g. hierarchical clustering methods). Visual Analytics systems need different algorithms for different phases of the work that can trade speed for quality in an understandable way.

Designing novel interaction and visualization techniques to explore huge data sets is an important goal and requires solving hard problems, but how can we assess whether or not our techniques and systems provide real improvements? Without this answer, we cannot know if we are heading in the right direction. This is why we have been actively involved in the design of evaluation methods for information visualization [63], [62], [60], [61], [58]. For more complex systems, other methods are required. For these we want to focus on longitudinal evaluation methods while still trying to improve controlled experiments.

3.2. Innovation

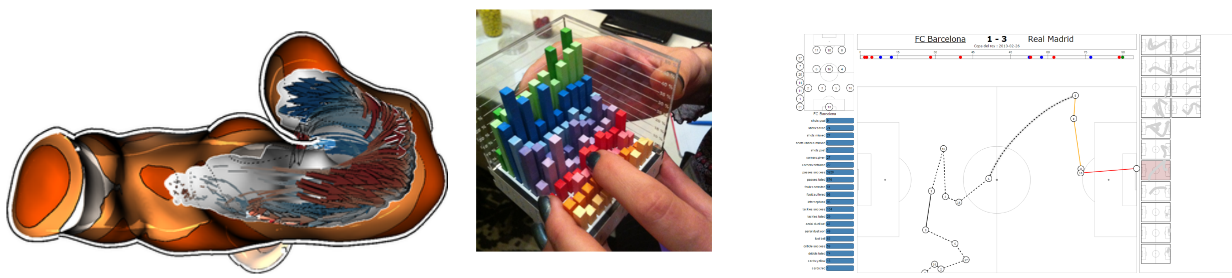


Figure 1. Example novel visualization techniques and tools developed by the team. Left: a non-photorealistic rendering technique that visualizes blood flow and vessel thickness. Middle: a physical visualization showing economic indicators for several countries, right: SoccerStories a tool for visualizing soccer games.

We design novel visualization and interaction techniques (see, for example, Figure 1). Many of these techniques are also evaluated throughout the course of their respective research projects. We cover application domains such as sports analysis, digital humanities, fluid simulations, and biology. A focus of Aviz' work is the improvement of graph visualization and interaction with graphs. We further develop individual techniques for the design of tabular visualizations and different types of data charts. Another focus is the use of animation as a transition aid between different views of the data. We are also interested in applying techniques from illustrative visualization to visual representations and applications in information visualization as well as scientific visualization [8].

3.3. Evaluation Methods

Evaluation methods are required to assess the effectiveness and usability of visualization and analysis methods. Aviz typically uses traditional HCI evaluation methods, either quantitative (measuring speed and errors) or qualitative (understanding users tasks and activities). Moreover, Aviz is also contributing to the improvement of evaluation methods by reporting on the best practices in the field, by co-organizing workshops (BELIV 2010–2018) to exchange on novel evaluation methods, by improving our ways of reporting, interpreting and communicating statistical results, and by applying novel methodologies, for example to assess visualization literacy [3], [4].

3.4. Software Infrastructures

We want to understand the requirements that software and hardware architectures should provide to support exploratory analysis of large amounts of data. So far, "big data" has been focusing on issues related to storage management and predictive analysis: applying a well-known set of operations on large amounts of data. Visual Analytics is about exploration of data, with sometimes little knowledge of its structure or properties. Therefore, interactive exploration and analysis is needed to build knowledge and apply appropriate analyses; this knowledge and appropriateness is supported by visualizations. However, applying analytical operations on large data implies long-lasting computations, incompatible with interactions, and generates large amounts of results, impossible to visualize directly without aggregation or sampling. Visual Analytics has started to tackle these problems for specific applications but not in a general manner, leading to fragmentation of results and difficulties to reuse techniques from one application to the other. We are interested in abstracting-out the issues and finding general architectural models, patterns, and frameworks to address the Visual Analytics challenge in more generic ways.

3.5. Emerging Technologies

Figure 2. Example emerging technology solutions developed by the team for multi-display environments, wall displays, and token-based visualization.

We want to use different types of display media to empower humans to visually and interactively explore information, in order to better understand and exploit it. This includes novel display equipment and accompanying input techniques. The Aviz team specifically focuses on the exploration of the use of large displays in visualization contexts as well as emerging physical and tangible visualizations (e. g. [6], [5]). In terms of interaction modalities our work focuses on using touch and tangible interaction. Aviz participates to the Digiscope project that funds 11 wall-size displays at multiple places in the Paris area (see http://www.digiscope.fr), connected by telepresence equipment and a Fablab for creating devices. Aviz is in charge of creating and managing the Fablab, uses it to create physical visualizations, and is also using the local wall-size display (called WILD) to explore visualization on large screens. The team also investigates the perceptual, motor and cognitive implications of using such technologies for visualization.

3.6. Psychology

More cross-fertilization is needed between psychology and information visualization. The only key difference lies in their ultimate objective: understanding the human mind vs. helping to develop better tools. We focus on understanding and using findings from psychology to inform new tools for information visualization. In many cases, our work also extends previous work in psychology. Our approach to the psychology of information visualization is largely holistic and helps bridge gaps between perception, action and cognition in the context of information visualization. Our focus includes the perception of charts in general, perception in large display environments, collaboration, perception of animations, how action can support perception and cognition, and judgment under uncertainty (e. g. [9]).

4. Application Domains

4.1. Natural Sciences

As part of a CORDI PhD project, we collaborate with researchers at CERN on interactive data visualization using augmented reality, with the goal to better understand this new visualization environment and to support the physicists in analysing their 3D particle collision data. As part of another CORDI PhD project, we collaborate with researchers at the German Center for Climate Computation (DKRZ), to better understand collaborative data exploration and interaction in immersive analytics contexts. Finally, as part of the Inria IPL "Naviscope," we collaborate with researchers at INRA (as well as other Inria teams) on interactive visualization tools for the exploration of plant embryo development.

4.2. Social Sciences

We collaborate with social science researchers from EHESS Paris on the visualization of dynamic networks; they use our systems (GeneaQuilts [56], Vistorian [64], PAOHVis [7]) and teach them to students and researchers. Our tools are used daily by ethnographers and historians to study the evolution of social relations over time. In the social sciences, many datasets are gathered by individual researchers to answer a specific question, and automated analytical methods cannot be applied to these small datasets. Furthermore, the studies are often focused on specific persons or organizations and not always on the modeling or prediction of the behavior of large populations. The tools we design to visualize complex multivariate dynamic networks are unique and suited to typical research questions shared by a large number of researchers. This line of research is supported by the DataIA "HistorIA" project, and by the "IVAN" European project. We also collaborate on the BART initiative, a joint project with IRT-SystemX on the analysis and visualization of blockchain data, in collaboration with economists from Université Paris-Sud.

4.3. Medicine

We collaborate with CMAP/Polytechnique on the analysis and visualization of CNAM Data "parcours de santé" to help referent doctors and epidemiologists make sense of French health data. In particular, we are working on a subset of the CNAM Data focused on urinary problems, and we have received a very positive feedback from doctors who can see what happens to the patients treated in France vs. what they thought happened through the literature. This project is starting but is already getting a lot of traction from our partners in medicine, epidemiology, and economy of health.

5. Highlights of the Year

5.1. Highlights of the Year

- Aviz researchers contributed 36 publications this year. Amongst these 3 papers were presented at IEEE VIS, the largest international Visualization and Visual Analytics conference.
- Aviz researchers remained active in the research community. We organized one Dagstuhl seminar and one Shonan seminar, served in 8 different organizing committee roles, on 19 program committees, and on 5 different journal editorial board. We reviewed for 25 different conference venues and for 12 different journals. We also gave 26 invited talks and served on two different steering committees and a best PhD award committee and served on 10 different juries (for both PhD and Master students).
- Aviz researchers started two four-year ANR grants as principle investigators.

5.1.1. Awards

- Tobias Isenberg was named the Associate Editor of the Year for Elsevier Computers & Graphic.
- Pierre Dragicevic and his co-authors received a best paper award at ACM CHI .
- Natkamon Tovanich and his co-authors received the Honorable Mention poster award at EuroVis 2019 .

BEST PAPERS AWARDS :

[34]
P. DRAGICEVIC, Y. JANSEN, A. SARMA, M. KAY, F. CHEVALIER.*Increasing the Transparency of Research Papers with Explorable Multiverse Analyses*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019 [*DOI :* 10.1145/3290605.3300295], https://hal.inria.fr/hal-01976951

[51]
N. TOVANICH, N. HEULOT, J.-D. FEKETE, P. ISENBERG.*A Systematic Review of Online Bitcoin Visualizations*, 2019, Posters of the European Conference on Visualization (EuroVis), The poster received the Honorable Mention award at EuroVis 2019 [*DOI :* 10.2312/EURP.20191148], https://hal.archives-ouvertes.fr/hal-02155171

6. New Software and Platforms

6.1. Cartolabe

KEYWORD: Information visualization

FUNCTIONAL DESCRIPTION: The goal of Cartolabe is to build a visual map representing the scientific activity of an institution/university/domain from published articles and reports. Using the HAL Database, Cartolabe provides the user with a map of the thematics, authors and articles . ML techniques are used for dimensionality reduction, cluster and topics identification, visualisation techniques are used for a scalable 2D representation of the results.

NEWS OF THE YEAR: Improvement of the graphical interface

- Participants: Philippe Caillou, Jean-Daniel Fekete, Jonas Renault and Anne-Catherine Letournel
- Partners: LRI - Laboratoire de Recherche en Informatique - CNRS
- Contact: Philippe Caillou
- URL: http://www.cartolabe.fr/

6.2. BitConduite

BitConduite Bitcoin explorer

KEYWORDS: Data visualization - Clustering - Financial analysis - Cryptocurrency

FUNCTIONAL DESCRIPTION: BitConduite is a web-based visual tool that allows for a high level explorative analysis of the Bitcoin blockchain. It offers a data transformation back end that gives us an entity-based access to the blockchain data and a visualization front end that supports a novel high-level view on transactions over time. In particular, it facilitates the exploration of activity through filtering and clustering interactions. This gives analysts a new perspective on the data stored on the blockchain.

- Authors: Jean-Daniel Fekete, Petra Isenberg and Christoph Kinkeldey
- Contact: Petra Isenberg

6.3. PAOHvis

Parallel Aggregated Ordered Hypergraph Visualization

KEYWORDS: Dynamic networks - Hypergraphs

FUNCTIONAL DESCRIPTION: Parallel Aggregated Ordered Hypergraph (PAOH) is a novel technique to visualize dynamic hypergraphs [26]. Hypergraphs are a generalization of graphs where edges can connect more than two vertices. Hypergraphs can be used to model co-authorship networks with multiple authors per article, or networks of business partners. A dynamic hypergraph evolves over discrete time slots. A PAOH display represents vertices as parallel horizontal bars and hyperedges as vertical lines that connect two or more vertices. We believe that PAOH is the first technique with a highly readable representation of dynamic hypergraphs without overlaps. It is easy to learn and is well suited for medium size dynamic hypergraph networks such as those commonly generated by digital humanities projects - our driving application domain (see Fig. 3).

- Contact: Paola Tatiana Llerena Valdivia
- URL: https://aviz.fr/paohvis

6.4. AR Collaborative Visualization

KEYWORDS: Augmented reality - Collaborative science - Android

FUNCTIONAL DESCRIPTION: Allows to look at VTK datasets using AR-HMD (Microsoft HoloLens) in multi-users environments (i.e., one headset per user). A Multi-touch tablet is provided per user to manipulate the environment.

- Contact: Mickael Sereno

6.5. Platforms

6.5.1. Vispubdata.org

AVIZ members are making available for research a dataset of IEEE VIS publications at http://vispubdata.org. This dataset is actively being used for research and conference organization.

7. New Results

7.1. A Model of Spatial Directness in Interactive Visualization

Participants: Stefan Bruckner [University of Bergen], Tobias Isenberg [correspondant], Timo Ropinski [University of Ulm], Alexander Wiebel [Hochschule Worms University of Applied Sciences].

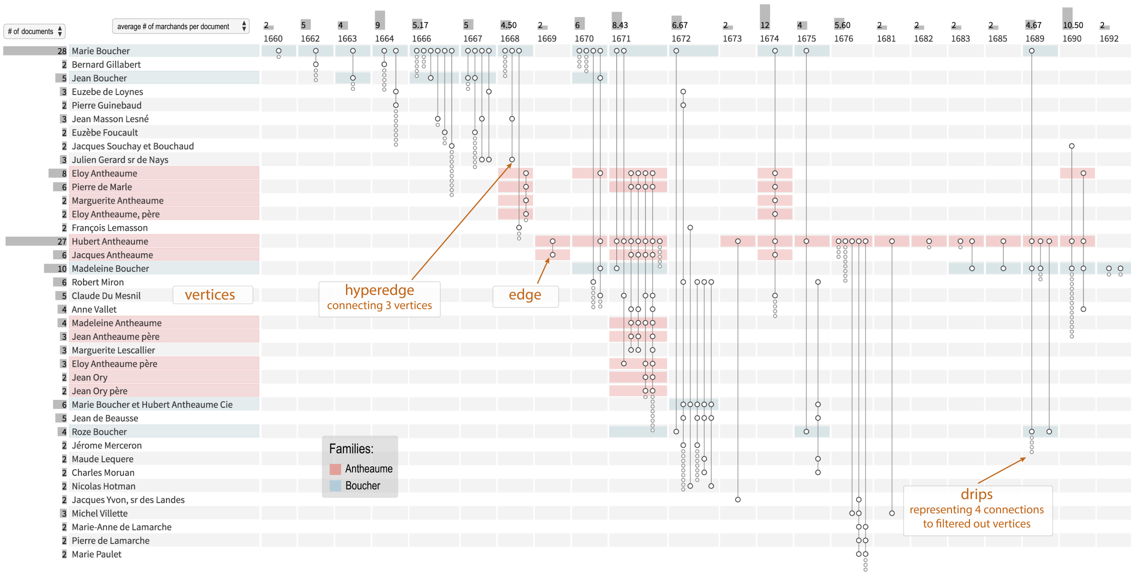


Figure 3. Using Dynamic Hypergraphs to Reveal the Evolution of the Business Network of a 17th Century French Woman Merchant

$$
\begin{tikzcd}
 & \mathscr{V} \arrow[r, "o"] \arrow[d, dashed, "vi"] & \mathscr{O} \arrow[dr, "u_p"] \arrow[d, dashed, "om"] & \\
\mathscr{D} \arrow[ur, "v"] & \mathscr{I} \arrow[l, "i"] & \mathscr{M} \arrow[l, "m"] & \mathscr{U} \arrow[l, "u_a"]
\end{tikzcd}
$$

Figure 4. Illustration of the model of interaction directness.

We discuss the concept of directness in the context of spatial interaction with visualization [2] . In particular, we propose a model that allows practitioners to analyze and describe the spatial directness of interaction techniques, ultimately to be able to better understand interaction issues that may affect usability. To reach these goals, we distinguish between different types of directness (Figure 4). Each type of directness depends on a particular mapping between different spaces, for which we consider the data space, the visualization space, the output space, the user space, the manipulation space, and the interaction space. In addition to the introduction of the model itself, we also show how to apply it to several real-world interaction scenarios in visualization, and thus discuss the resulting types of spatial directness, without recommending either more direct or more indirect interaction techniques. In particular, we will demonstrate descriptive and evaluative usage of the proposed model, and also briefly discuss its generative usage.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Bruckner2019MSD.

7.2. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses

Participants: Pierre Dragicevic [correspondant], Yvonne Jansen [CNRS], Abhraneel Sarma [University of Michigan], Matthew Kay [University of Michigan], Fanny Chevalier [University of Toronto].

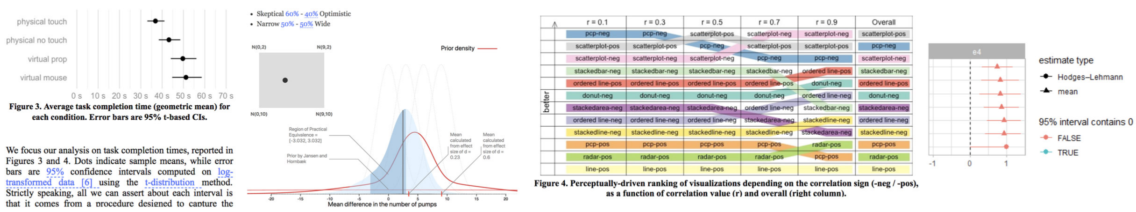


Figure 5. Examples of explorable multiverse analyses.

We presented explorable multiverse analysis reports, a new approach to statistical reporting where readers of research papers can explore alternative analysis options by interacting with the paper itself [34]. This approach draws from two recent ideas: i) multiverse analysis, a philosophy of statistical reporting where paper authors report the outcomes of many different statistical analyses in order to show how fragile or robust their findings are; and ii) explorable explanations, narratives that can be read as normal explanations but where the reader can also become active by dynamically changing some elements of the explanation. Based on five examples and a design space analysis, we showed how combining those two ideas can complement existing reporting approaches and constitute a step towards more transparent research papers. This work received a best paper award at ACM CHI.

More on the project Web page, including interactive demos: https://explorablemultiverse.github.io/.

7.3. Glanceable Visualizations for Smartwatches

Participants: Tanja Blascheck [correspondant], Lonni Besançon [Linköping University], Anastasia Bezerianos, Bongshin Lee [Microsoft Research], Petra Isenberg.

The goal of this project is to study very small data visualizations, micro visualizations, in display contexts that can only dedicate minimal rendering space for data representations. Specifically, we define micro visualizations as small-scale visualizations that lack or have a limited set of reference structures such as labels, data axes, or grid lines and have a small physical footprint of a few square centimeters. Micro visualizations can be as simple as small unit-based visualizations such as a battery indicator but also include multi-dimensional visualizations such as star glyphs, small geographic visualizations or even small network visualizations.

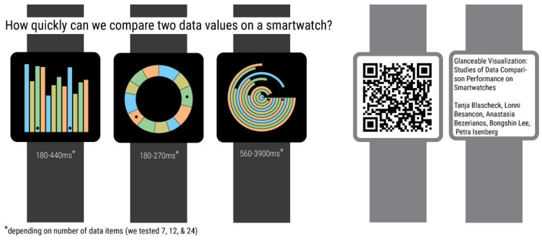


Figure 6. Comparison of bar, donut, and radial bar charts on a smartwatch.

Although micro visualizations are essential to mobile visualization contexts, we know surprisingly little about their general visual and interaction design space or people's ability in interpreting micro visualizations. We will address this gap by proposing a common framework, conducting empirical studies to understand people's abilities to interpret these visualizations while in motion, and by developing a software toolkit to aid practitioners in developing micro visualizations for emerging mobile and wearable displays.

In summary, we aim at paving the way for a pervasive use of visualizations and thus a better and broader understanding of the complex world around us.

More information in related publications ([1],[48]) and on the project Web page: https://www.aviz.fr/smartwatchperception.

7.4. Hybrid Touch/Tangible Spatial 3D Data Selection

Participants: Lonni Besançon [Linköpping University], Mickael Sereno [correspondant], Lingyun Yu [Hangzhou Dianzi University], Mehdi Ammi [University of Paris 8], Tobias Isenberg.

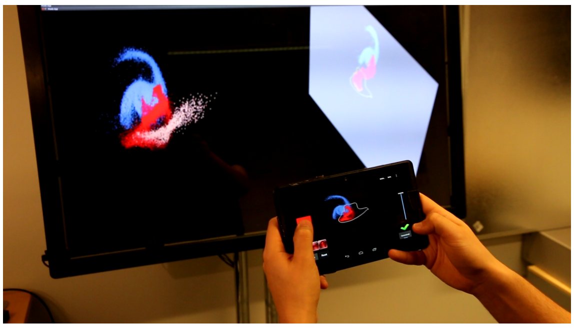


Figure 7. Illustration of Tangible Brush application which combines a spatial-aware multi-touch tablet and a remote large screen which shows different perspectives of the view shared with the tablet.

We discussed spatial selection techniques for three-dimensional datasets. Such 3D spatial selection is fundamental to exploratory data analysis. While 2D selection is efficient for datasets with explicit shapes and structures, it is less efficient for data without such properties.

We first proposed a new taxonomy of 3D selection techniques [12], focusing on the amount of control the user has to define the selection volume. We then described the 3D spatial selection technique Tangible Brush (Figure 7), which gives manual control over the final selection volume. It combines 2D touch with 6-DOF 3D tangible input to allow users to perform 3D selections in volumetric data. We use touch input to draw a 2D lasso, extruding it to a 3D selection volume based on the motion of a tangible, spatially-aware tablet. We described our approach and presented its quantitative and qualitative comparison to state-of-the-art structure-dependent selection. Our results show that, in addition to being dataset-independent, Tangible Brush is more accurate than existing dataset-dependent techniques, thus providing a trade-off between precision and effort.

7.5. Is there a reproducibility crisis around here? Maybe not, but we still need to change

Participants: Alex Holcombe [The University of Sydney], Charles Ludowici [The University of Sydney], Steve Haroz.

Those of us who study large effects may believe ourselves to be unaffected by the reproducibility problems that plague other areas [39]. However, we will argue that initiatives to address the reproducibility crisis, such as preregistration and data sharing, are worth adopting even under optimistic scenarios of high rates of replication success. We searched the text of articles published in the Journal of Vision from January through October of 2018 for URLs (our code is here: https://osf.io/cv6ed/) and examined them for raw data, experiment code, analysis code, and preregistrations. We also reviewed the articles' supplemental material. Of the 165 articles, approximately 12% provide raw data, 4% provide experiment code, and 5% provide analysis code. Only one article contained a preregistration. When feasible, preregistration is important because p-values are not interpretable unless the number of comparisons performed is known, and selective reporting appears to be common across fields. In the absence of preregistration, then, and in the context of the low rates of successful replication found across multiple fields, many claims in vision science are shrouded by uncertain credence. Sharing de-identified data, experiment code, and data analysis code not only increases credibility and ameliorates the negative impact of errors, it also accelerates science. Open practices allow researchers to build on others' work more quickly and with more confidence. Given our results and the broader context of concern by funders, evident in the recent NSF statement that "transparency is a necessary condition when designing scientifically valid research" and "pre-registration..." can help ensure the integrity and transparency of the proposed research", there is much to discuss.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Participants: Yuheng Feng, Jean-Daniel Fekete, Alejandro Ribs. Project title: *Visual Sensitivity Analysis for Ensembles of Curves*: The goal of this project is to investigates new progressive methods to compute PCA over large amounts of time-series in interactive time.

9. Partnerships and Cooperations

9.1. National Initiatives

- PCR ANR project EMBER "Situated Visualizations for Personal Analytics". Duration: 48 months. Total funding: 712 k€. Partners: Inria Saclay, Inria Bordeaux, Sorbonne Université. Coordinator: Pierre Dragicevic. See website: http://ember.inria.fr/.
- Naviscope Inria Project Lab on Image-guided NAvigation and VIsualization of large data sets in live cell imaging and microSCOPy; collaboration with several Inria project teams and external collaborators; this grant supports a PhD position and funds travel and equipment.

9.2. European Initiatives

9.2.1. Collaborations in European Programs, Except FP7 & H2020

Program: ANR PRCI

Project acronym: MicroVis

Project title: Micro visualizations for pervasive and mobile data exploration

Duration: 11/2019 - 08/2022

Coordinator: Petra Isenberg

Other partners: University of Stuttgart

Abstract: The goal of this joint Franco-German project is to study very small data visualizations, micro visualizations, in display contexts that can only dedicate minimal rendering space for data representations. We will study human perception of and interaction with micro visualizations given small as well as complex data. The increasing demand for data visualizations on small mobile devices such as fitness tracking armbands, smart watches, or mobile phones drives our research. Given this usage context, we focus on situations in which visualizations are used "on the go," while walking, riding a vehicle, or running. It is still unclear to which extent our knowledge of desktop-sized visualizations transfers to contexts that involve minimal display space, diverse viewing angles, and moving displays.

Program: 2016 FWF–ANR Call for French-Austrian Joint Projects

Project acronym: ILLUSTRARE

Project title: Integrative Visual Abstraction of Molecular Data

Duration: 48 months

Coordinator: Tobias Isenberg and Ivan Viola

Other partners: TU Wien, Austria

Abstract: The essential building block of visualization is the phenomenon of visual abstraction. While visual abstraction is intuitively understood, there is no scientific theory associated with it that would be useful in the visualization synthesis process. Our central aim of this project is thus to gain better understanding of the visual abstraction characteristics. We lay down a hypothetical initial basis of theoretical foundations of visual abstractions in the proposal. We hypothesize that visual abstraction is a multidimensional phenomenon that can be spanned by axes of abstraction. Besides abstractions associated with a static structure we take a closer look at abstractions related to dynamics, procedures, and emergence of the structure. We also study abstraction characteristics related to multi-scale phenomena defined both in space and in time. This hypothetical basis is either supported or rejected by means of exemplary evidence from the specific application domain of structural biology. Structural biology data is very complex, it includes the aspect of emergence and it is defined over multiple scales. Furthermore, abstraction has led to key discoveries in biology, such as the organization of the DNA. We study the multiscale visual abstraction characteristics on the visualization of long nucleic strands and the abstractions that convey emerging phenomena on visualization of molecular machinery use cases. From these two fields we work toward a theory of visual abstraction in a bottom-up manner, investigating the validity of the theory in other application domains as well.

Program: CHIST-ERA

Project acronym: IVAN

Project title: Interactive and Visual Analysis of Networks

Duration: May 2018 - April 2021

Coordinator: Dr. Torsten Möller, Uni Wien, Austria

Other partners: EPFL, Switzerland, Inria France, Uni Wien, Austria

Abstract: The main goal of IVAN is to create a visual analysis system for the exploration of dynamic or time-dependent networks (from small to large scale). Our contributions will be in three principal areas:

1. novel algorithms for network clustering that are based on graph harmonic analysis and level-of-detail methods;
2. the development of novel similarity measures for networks and network clusters for the purpose of comparing multiple network clusterings and the grouping (clustering) of different network clusterings; and
3. a system for user-driven analysis of network clusterings supported by novel visual encodings and interaction techniques suitable for exploring dynamic networks and their clusterings in the presence of uncertainties due to noise and uncontrolled variations of network properties.

Our aim is to make these novel algorithms accessible to a broad range of users and researchers to enable reliable and informed decisions based on the network analysis.

9.2.2. Collaborations with Major European Organizations

The Bauhaus-Universität Weimar (Germany)

Steve Haroz collaborates with Florian Echtler to analyze research transparency in human-computer interaction.

Hasso Plattner Institute (Germany)

Pierre Dragicevic and Tobias Isenberg collaborate with Amir Semmo on stylization filters for facilitating the examination of disturbing visual content.

University of Zurich (Switzerland)

Pierre Dragicevic and Steve Haroz collaborate with Chat Wacharamanotham on transparent statistical reporting and efficient statistical communication.

KU Leuven (Belgium)

Pierre Dragicevic collaborates with Andrew Vande Moere on a survey on data physicalization.

Linköping University (Sweden)

Tobias Isenberg, Xiyao Wang, and Mickael Sereno collaborate with Lonni Besançon on interaction with 3D visualization.

University of Granada (Spain)

Tobias Isenberg collaborates with Domingo Martin and German Arroyo on digital stippling.

University of Roma (Italy), TU Darmstadt (Germany)

Jean-Daniel Fekete Fekete collaborates with Giuseppe Santucci, Carsten Binnig and colleagues on the design of database benchmarks to better support visualization;

University of Bari (Italy)

Jean-Daniel Fekete collaborates with Paolo Buono on hypergraph visualization;

University of Konstanz (Germany)

Petra Isenberg collaborated with Johannes Fuchs and Anastasia Bezerianos on visualization for teaching clustering algorithms.

9.3. International Initiatives

9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

9.3.1.1. SEVEN

Title: Situated and Embedded Visualization for Data Analysis

International Partner (Institution - Laboratory - Researcher):

University of Calgary (Canada) - ILab - Wesley Willett

Start year: 2018

See also: http://aviz.fr/seven

The goal of this joint work between the Aviz team at Inria Saclay and the ILab at the University of Calgary is to develop and study situated data visualizations to address the limitations of traditional platforms of data analytics. In a situated data visualization, the data is directly visualized next to the physical space, object, or person it refers to. Situated data visualizations can surface information in the physical environment and allow viewers to interpret data in-context, monitor changes over time, make decisions, and act on the physical world in response to the insights gained. However, research on this topic remains scarce and limited in scope. We will build on our track record of successfull collaborations to jointly develop situated visualization as a novel research direction. The objective for the first year is to design and implement situated visualizations to support health and aging. Our joint work is expected to generate benefits at multiple levels, including to society and industry (by empowering individuals and professionals with technology), to the scientific community (by developing a new research direction), to the academic partners (by reinforcing existing research links and establishing them as leaders on the topic), and to students (by providing them with unique training opportunities with a diverse team of world-class researchers).

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

Microsoft Research: Petra Isenberg, Tobias Isenberg, and Tanja Blascheck regularly collaborate with Bong-shin Lee on topics related to non-desktop visualizations such as mobile visualization, ubiquitous visualization, or touch interaction for visualization.

University of Maryland: Catherine Plaisant regularly collaborates with various team members on projects related to temporal exploratory visualization.

9.3.3. Participation in Other International Programs

9.3.3.1. Inria International Chairs

IIC PLAISANT Catherine

Title: Visual Analytics for Exploratory Data Analysis

International Partner (Institution - Laboratory - Researcher):

University of Maryland (United States) - HCIL - Catherine Plaisant

Duration: 2018 - 2022

Start year: 2018

Visual Analytics for Exploratory Data Analysis: The project leverages Dr. Plaisants 30 years of experience in the design and evaluation of novel user interface and the longstanding synergies between my research activities and those of the AVIZ lab. It also builds on early collaborative activities having taken place between Maryland and Inria during a 2017 summer visit. The joint work particularly focuses on: event analysis, network analysis, and novel evaluation methods for visual analytics.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Catherine Plaisant (June–July): Invited professor from the University of Maryland, USA. Invited through a DigiCosme grant, Catherine Plaisant has spent two months with Aviz. We have launched two research projects, one on hypergraph visualization and one on tracing users to understand their use of visualization. Catherine Plaisant has interacted with all of the Aviz students and post-doctoral fellows, as well as with the permanent researchers.
- Paolo Buono, from the University of Bari, Italy (August–September): Paolo Buono has spent two months with Aviz working on the visualization of dynamic networks. He has collaborated with Paoa Valdivia, Catherine Plaisant, and Jean-Daniel Fekete for that project. He has also interacted with all the members of Aviz.
- Claudio Silva (August 2018 – June 2019): Sabbatical from New York University (USA). Also, invited professor through a DigiCosme grant for 3 months. Claudio Silva is spending one year with Aviz. We launched a bi-weekly seminar on explainable machine-learning with visualization.
- Wesley Willett and Lora Oehlberg (June): as part of the associated team SEVEN both professors came for a three-day workshop to Aviz during which we discussed designs for the noise project sensors and the associated data displays. We also worked in more depth on a survey article we plan to publish.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Steve Haroz is a co-founder and co-organizer of Transparent Statistics in Human-Computer Interaction.
- Steve Haroz co-organized the Visualization for Communication 2019 workshop (http://viscomm.io).
- Petra Isenberg co-organized a Dagstuhl seminar on Mobile Data Visualization (https://www.dagstuhl.de/19292).
- Petra Isenberg co-organized a tutorial at the Ubicomp conference about Visualization for Ubicomp.
- Jean-Daniel Fekete co-organized a Shonan seminar on Interactive Visualization for Interpretable Machine Learning (https://shonan.nii.ac.jp/seminars/161/).

10.1.1.2. Member of the Organizing Committees

- Steve Haroz was Open Practices chair of the IEEE Information Visualization conference (until July 2019).
- Jean-Daniel Fekete was a member of the VGTC Best Thesis Award 2019.
- Tanja Blascheck was Video and Demo co-chair for ETRA 2019.
- Tanja Blascheck was Publicity chair for VISSOFT 2019.
- Petra Isenberg is a member of the IEEE reVISe committee 2019–2020.
- Petra Isenberg is the ACM SigCHI Paris Treasurer (2017–2019)
- Jean-Daniel Fekete was associate editor-in-chief of TVCG, liaison with the IEEE VIS paper chairs.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Petra Isenberg was paper chair for IEEE Information Visualization

10.1.2.2. Member of the Conference Program Committees

- Jean-Daniel Fekete was a member of the program committee for VIS.
- Petra Isenberg was a member of the ACM CHI program committee.
- Tobias Isenberg was a member of the program committee for IEEE SciVis, IVAPP, ACM/Eurographics Expressive, TrustVis Workshop, EuroVis, ACM IUI, IEEE VR.
- Pierre Dragicevic was a member of the program committee for InfoVis.
- Paola Valdivia was a member of the program committee for Short Papers of VIS.
- Tanja Blascheck was a member of the program committee for VMV, LEVIA, VISSOFT, EuroVis Posters, EuroVis Short Papers, ETRA, ETVIS, VAST.

10.1.2.3. Reviewer

- Steve Haroz reviewed for CHI, VIS
- Jean-Daniel Fekete reviewed for CHI
- Tobias Isenberg reviewed for EuroVis, ACM/Eurographics Expressive, ACM IUI, ACM SIGGRAPH, IEEE PacificVis, TrustVis, ACM UIST, IEEE VIS, IEEE VR
- Pierre Dragicevic reviewed for CHI, VIS, Interact, IHM
- Tanja Blascheck reviewed for LEVIA, VMV, VISSOFT, VAST, EuroVis, PacificVis, ETRA, ETVIS, CHI, EuroVA, PacificVAST, VAST Challenge, Vision X Vis, Vis Short Papers
- Paola Valdivia reviewed for Short Papers of VIS 2019
- Xiyao Wang reviewed for CHI, VR, ICMI, MobileHCI, Vis
- Petra Isenberg reviewed for CHI, EuroVis

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Pierre Dragicevic is member of the editorial board of the Journal of Perceptual Imaging (JPI) and the Springer Human–Computer Interaction Series (HCIS).
- Jean-Daniel Fekete is associate editor-in-chief of IEEE Transactions on Visualization and Computer Graphics.
- Petra Isenberg is associate editor of IEEE Transactions on Visualization and Computer Graphics.
- Petra Isenberg is associate editor-in-chief at IEEE Computer Graphics & Applications.
- Tobias Isenberg is member of the editorial board of Elsevier's Computers & Graphics journal.

10.1.3.2. Reviewer - Reviewing Activities

- Steve Haroz reviewed for Cognition, Advances in Methods and Practices in Psychological Science, and Meta-Psychology
- Jean-Daniel Fekete reviewed for TVCG
- Tobias Isenberg reviewed for C&G, TIIS, TVCG
- Pierre Dragicevic reviewed for TVCG, CG&A, IV, Behavior & Information Technology, JPI
- Tanja Blascheck reviewed for IWC, TOCCHI, TVCG
- Petra Isenberg reviewed for TVCG

10.1.4. Invited Talks

- Steve Haroz: “How the Brain Perceives and Remembers Quantities“. Graphics Hunters Show Conference - Utrecht, The Netherlands.
- Steve Haroz: “How the Brain Perceives and Remembers Quantities“. Ecole Normal Superior. Paris, France.
- Steve Haroz: “How the Brain Perceives and Remembers Quantities“. Keynote at VisCrowd 2019. Clermont-Ferrand, France.

- Steve Haroz: “How the Brain Perceives and Remembers Quantities“. Datarama 2019. Nantes, France.
- Steve Haroz: “Set Comparison Is Imprecise and Prone to Bias“. VisXVision workshop at IEEE VIS 2019. Vancouver, Canada.
- Steve Haroz: “Transparency in Empirical Research“. Panel talk at IEEE VIS 2019. Vancouver, Canada.
- Steve Haroz: “Using Open Science to Improve Credibility in Visualization, HCI, and Vision Science Research“. SFB/TRR 161. Waldachtal, Germany.
- Jean-Daniel Fekete: ”Practical Use Cases for Progressive Visual Analytics”, Seminar at Shandong University, Qingdao, China, Nov. 26, 2019
- Jean-Daniel Fekete: ”Dynamic Social Networks and Progressive Data Analysis”, Seminar at Ochanomizu University, Tokyo, Japan, Nov. 22, 2019
- Jean-Daniel Fekete: ”Advances in Network Visualization”, Keynote at MARAMI 2019, Dijon, France, Nov. 8, 2019
- Jean-Daniel Fekete: ”Multidimensional Projection at Scale on the Web: Tips and Tricks”, Keynote of the HILDA 2019 Workshop, Amsterdam, The Netherlands, July 5th, 2019.
- Jean-Daniel Fekete: ”Exploration visuelle d’hypergraphes dynamiques”, Seminar at INED, Paris, France, June 18th, 2019.
- Jean-Daniel Fekete: ”Données, humanité(s), démarche scientifique”, Panel at Inria Scientific Days 2019, Lyon, France, June 6th, 2019.
- Jean-Daniel Fekete: ”Analyzing the Evolution of Relationships with Dynamic Hypergraphs”, Invited Talk, University of Delft, Delft, The Netherlands, Apr. 8th, 2019
- Jean-Daniel Fekete: ”Analyzing the Evolution of Relationships with Dynamic Hypergraphs”, Invited Talk, EHESS, Paris, France, Apr. 3, 2019
- Jean-Daniel Fekete: ”Multidimensional Projection at Scale on the Web: Tips and Tricks”, Invited Talk, VISUS Group, Univ. of Stuttgart, Stuttgart, Germany, Mar. 1, 2019
- Tobias Isenberg: “Interactive 3D Data Exploration”, TU Dresden, Germany, Dec, 6, 2019
- Tobias Isenberg and Xiyao Wang: “VR and AR Visualizations in High-Energy Physics”, Learning to Discover: Advanced Pattern Recognition Workshop, Institut Pascal, Orsay, France, Oct. 18, 2019
- Tobias Isenberg: “Carnivorous Plants and How to Visualize their Habitats based on Social Media Data”, VieVisDays, TU Wien, Austria, Aug. 20, 2019
- Tobias Isenberg and Xiyao Wang: “Understanding TrackML Results with a Visualization System using a PC + HoloLens Hybrid”, TrackML workshop at CERN, Geneva, Switzerland, Jul. 1, 2019
- Tobias Isenberg: “Illustrative Visualization and Interactive Exploration of Three-Dimensional Scientific Data”, Synchrotron SOLEIL, France, May. 20, 2019
- Tobias Isenberg: “Illustrative Visualization and Interactive Exploration of Three-Dimensional Scientific Data”, Shandong University, Qingdao, China, Feb. 24, 2019
- Tanja Blascheck: “Understanding Interactive Visualizations”, Inria Sophia-Antipolis, Nice, France, January 2019.
- Petra Isenberg: “Top Research Questions for Empirical Studies in Visualization”, Panel statement given at IEEE VIS 2019, Vancouver, Canada.
- Petra Isenberg: “Physical, Contextual, and Full of Value? What do novel directions in Visualization teach us about judging the value of visualization?”, GiCentre Seminar Talk, City, University of London, UK
- Petra Isenberg: “Physical, Contextual, and Full of Value? What do novel directions in Visualization teach us about judging the value of visualization?”, Keynote at BioVis 2019 at the ISMB conference in Basel, Switzerland

10.1.5. Leadership within the Scientific Community

- Jean-Daniel Fekete was the chair of the EuroVis 2019 Best PhD Award.
- Petra Isenberg and Tobias Isenberg are members of the International Beliv workshop steering committee.
- Tobias Isenberg is a member of the steering comittee for ACM/Eurographics Expressive Graphics.

10.1.6. Research Administration

- Petra Isenberg was vice-president of the Inria Saclay CR hiring committee
- Pierre Dragicevic is Member of the Commission Consultative de Spécialistes de l'Université Paris-Sud (CCSU).
- Pierre Dragicevic is co-chair of the DataSense Axis of the Labex DigiCosme.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : Petra Isenberg, Interactive Information Visualization, 21h en equivalént, niveau (M1, M2), Univeristé Paris Sud, France.

Master : Petra Isenberg, Graphisme et Visualisation, 24h en equivalént TD, niveau (M2), Polytech Paris-Sud, France.

Master : Petra Isenberg, Visual Analytics, 48 equivalént, niveau (M2), CentraleSupelec, France.

Licence : Tobias Isenberg, "Introduction to Computer Graphics", 12h en équivalent TD, L3, Polytech Paris-Sud, France.

Master : Tobias Isenberg, "Photorealistic Rendering/Advanced Computer Graphics", 21h en équivalent TD, M1, Polytech Paris-Sud and Université Paris-Saclay, France.

10.2.2. Supervision

PhD in progress: Alexis Pister, *Exploration, analyse, interprétabilité de larges réseaux historiques*, Université Paris-Saclay, defense planned for 2022, Jean-Daniel Fekete, Christophe Prieur, Inria.

PhD in progress: Yuheng Feng, *Visualisation pour l'analyse progressive"*, Université Paris-Saclay, defense planned for 2022, Jean-Daniel Fekete, Université Paris-Saclay

PhD in progress: Mickaël Sereno, *Collaborative Data Exploration and Discussion Supported by AR*, Univ. Paris-Sud, defense planned for September 2021, Tobias Isenberg

PhD in progress: Xiyao Wang, *Augmented Reality Environments for the Interactive Exploration of 3D Data*, Univ. Paris-Sud, defense planned for October 2020, Tobias Isenberg

PhD in progress: Sarkis Halladjian, *Spatially Integrated Abstraction of Genetic Molecules*, Univ. Paris-Sud; defense planned for September 2020, Tobias Isenberg

PhD: Haichao Miao, *Geometric Abstraction for Effective Visualization and Modeling*, TU Wien, Austria, August 5, 2019, supervised by Ivan Viola, co-supervised by Ivan Barišic, Tobias Isenberg, and Eduard Gröller.

PhD in progress: Mohamad Alaul Islam: *MicroVisualization for mobile and ubiquitous data exploration*, Université Paris-Saclay, defense planned for 2022, Petra Isenberg and Jean-Daniel Fekete.

PhD in progress: Natkamon Tovanich: *Blockchain Visual Analytics*, defense planned for 2021, Université Paris-Saclay, Petra Isenberg and Jean-Daniel Fekete.

10.2.3. Juries

- Pierre Dragicevic: PhD proposal evaluation committee of Luiz Morais (Universidade Federal de Campina Grande, Brésil).
- Pierre Dragicevic: Reviewer for Hu Xi's M2 internship.

- Pierre Dragicevic: Co-Supervisor for Yumin Hong's M2 internship.
- Pierre Dragicevic: Reviewer for Talent Doctoral Fellowship Programme at the University of Copenhagen.
- Pierre Dragicevic: Reviewer for ATER 2019 applications at the Université Paris Saclay.
- Petra Isenberg: Reviewer for GRIOUI Fairouz M2 internship.
- Jean-Daniel Fekete: Member of the PhD committee of Dr. Philip Tchernavskij
- Jean-Daniel Fekete: Member of the PhD committee of Dr. Hugo Romat.
- Jean-Daniel Fekete: Member of the PhD committee of Dr. Sriram Karthik Badam.
- Jean-Daniel Fekete: Member of the PhD committee of Dr. Nicola Pezzotti.

10.3. Popularization

10.3.1. Articles and contents

- Steve Haroz's twitter account on a broad range of topics including open science, visualization, and visual perception have been viewed over 1.1 million times and received over 30,000 engagements in the form of clicks, likes, and retweets in 2019.
- Steve Haroz maintains a website cataloging openly accessible papers and research materials in visualization research. Open Access Vis has averaged 2,000 unique visitors per year for the past three years.
- Steve Haroz maintains a research website receiving 5,000 unique visitors and a blog on research receiving 2,000 unique visitors in 2019.
- Pierre Dragicevic and Yvonne Jansen: the data physicalization wiki and the List of Physical Visualizations and Related Artefacts (600 weekly visits) are continuously being updated.
- Matthew Brehmer wrote a Medium article "Visualizing Trends on Mobile Phones: Animation or Small Multiples?" on an Aviz co-authored publication https://medium.com/multiple-views-visualization-research-explained/mobiletrendvis-a948fa65bccd

10.3.2. Interventions

- Lonni Besançon gave a TEDx talk on work conducted at Aviz with Pierre Dragicevic and Tobias Isenberg [11]. Watch the talk here: https://www.youtube.com/watch?v=pDHomZ8FEoU.
- Jean-Daniel Fekete: Panel at Imaginascience 2019, « Faire parler les données : le défi de la data visualisation », Annecy, October 16, 2019

10.3.3. Creation of media or tools for science outreach

Petra Isenberg collaborated on research to develop a visualization tool to better understand clustering algorithms. The tool is available at https://educlust.dbvis.de/.

11. Bibliography

Major publications by the team in recent years

[1] T. Blascheck, L. Besançon, A. Bezerianos, B. Lee, P. Isenberg.*Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches*, in "IEEE Transactions on Visualization and Computer Graphics", January 2019, vol. 25, n^{o} 1, p. 616–629 [*DOI :* 10.1109/TVCG.2018.2865142], https://hal.inria.fr/hal-01851306

[2] S. BRUCKNER, T. ISENBERG, T. ROPINSKI, A. WIEBEL.*A Model of Spatial Directness in Interactive Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", August 2019, vol. 25, n^o 8, p. 2514-2528 [*DOI :* 10.1109/TVCG.2018.2848906], https://hal.inria.fr/hal-01813889

[3] S. HAROZ.*Open Practices in Visualization Research*, in "Evaluation and Beyond - Methodological Approaches for Visualization", Berlin, Germany, October 2018 [*DOI :* 10.31219/OSF.IO/8AG3W], https://hal.inria.fr/hal-01947432

[4] T. ISENBERG, P. ISENBERG, J. CHEN, M. SEDLMAIR, T. MÖLLER.*A Systematic Review on the Practice of Evaluating Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", 2013, vol. 19, n^o 12, p. 2818-2827 [*DOI :* 10.1109/TVCG.2013.126], https://hal.inria.fr/hal-00846775

[5] Y. JANSEN, P. DRAGICEVIC, P. ISENBERG, J. ALEXANDER, A. KARNIK, J. KILDAL, S. SUBRAMANIAN, K. HORNBÆK.*Opportunities and Challenges for Data Physicalization*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)", New York, NY, United States, ACM, April 2015 [*DOI :* 10.1145/2702123.2702180], https://hal.inria.fr/hal-01120152

[6] M. LE GOC, L. H. KIM, A. PARSAEI, J.-D. FEKETE, P. DRAGICEVIC, S. FOLLMER.*Zooids: Building Blocks for Swarm User Interfaces*, in "Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST)", Tokyo, Japan, October 2016 [*DOI :* 10.1145/2984511.2984547], https://hal.inria.fr/hal-01391281

[7] P. R. VALDIVIA, P. BUONO, C. PLAISANT, N. DUFOURNAUD, J.-D. FEKETE.*Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", 2020 [*DOI :* 10.1109/TVCG.2019.2933196], https://hal.inria.fr/hal-02264960

[8] I. VIOLA, T. ISENBERG.*Pondering the Concept of Abstraction in (Illustrative) Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", September 2018, vol. 24, n^o 9, p. 2573–2588 [*DOI :* 10.1109/TVCG.2017.2747545], https://hal.inria.fr/hal-01581177

[9] L. YUAN, S. HAROZ, S. FRANCONERI.*Perceptual proxies for extracting averages in data visualizations*, in "Psychonomic Bulletin and Review", April 2019, vol. 26, n^o 2, p. 669-676 [*DOI :* 10.3758/S13423-018-1525-7], https://hal.inria.fr/hal-02189581

[10] E. ZGRAGGEN, A. GALAKATOS, A. CROTTY, J.-D. FEKETE, T. KRASKA.*How Progressive Visualizations Affect Exploratory Analysis*, in "IEEE Transactions on Visualization and Computer Graphics", 2017, vol. 23, n^o 8, p. 1977-1987 [*DOI :* 10.1109/TVCG.2016.2607714], https://hal.inria.fr/hal-01377896

Publications of the year

Articles in International Peer-Reviewed Journal

[11] L. BESANÇON, A. SEMMO, D. J. BIAU, B. FRACHET, V. PINEAU, E. H. SARIALI, M. SOUBEYRAND, R. TAOUACHI, T. ISENBERG, P. DRAGICEVIC.*Reducing Affective Responses to Surgical Images and Videos Through Stylization*, in "Computer Graphics Forum", 2020, vol. 39, To appear, forthcoming [*DOI :* 10.1111/CGF.13886], https://hal.inria.fr/hal-02381513

[12] L. BESANÇON, M. SERENO, L. YU, M. AMMI, T. ISENBERG.*Hybrid Touch/Tangible Spatial 3D Data Selection*, in "Computer Graphics Forum", June 2019, vol. 38, n^o 3, p. 553-567 [*DOI :* 10.1111/CGF.13710], https://hal.inria.fr/hal-02079308

[13] T. Blascheck, L. Besançon, A. Bezerianos, B. Lee, P. Isenberg.*Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches*, in "IEEE Transactions on Visualization and Computer Graphics", January 2019, vol. 25, n^{o} 1, p. 616–629 [*DOI :* 10.1109/TVCG.2018.2865142], https://hal.inria.fr/hal-01851306

[14] M. Brehmer, B. Lee, P. Isenberg, E. K. Choe.*Visualizing Ranges over Time on Mobile Phones: A Task-Based Crowdsourced Evaluation*, in "IEEE Transactions on Visualization and Computer Graphics", 2019, vol. 25, n^{o} 1, p. 619-629 [*DOI :* 10.1109/TVCG.2018.2865234], https://hal.inria.fr/hal-01857469

[15] M. Brehmer, B. Lee, P. Isenberg, E. K. Choe.*A Comparative Evaluation of Animation and Small Multiples for Trend Visualization on Mobile Phones*, in "IEEE Transactions on Visualization and Computer Graphics", 2020, vol. 26, n^{o} 1, p. 364–374 [*DOI :* 10.1109/TVCG.2019.2934397], https://hal.inria.fr/hal-02317687

[16] S. Bruckner, T. Isenberg, T. Ropinski, A. Wiebel.*A Model of Spatial Directness in Interactive Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", August 2019, vol. 25, n^{o} 8, p. 2514-2528 [*DOI :* 10.1109/TVCG.2018.2848906], https://hal.inria.fr/hal-01813889

[17] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, P. Dragicevic.*A Task-based Taxonomy of Cognitive Biases for Information Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", 2019, forthcoming [*DOI :* 10.1109/TVCG.2018.2872577], https://hal.sorbonne-universite.fr/hal-01868738

[18] J.-D. Fekete, T. J. Jankun-Kelly, M. Tory, K. Xu.*Provenance Analysis for Sensemaking*, in "IEEE Computer Graphics and Applications", November 2019, vol. 39, n^{o} 6, p. 27-29 [*DOI :* 10.1109/MCG.2019.2945378], https://hal.inria.fr/hal-02344205

[19] S. Halladjian, H. Miao, D. Kouřil, E. Gröller, I. Viola, T. Isenberg.*ScaleTrotter: Illustrative Visual Travels Across Negative Scales*, in "IEEE Transactions on Visualization and Computer Graphics", January 2020, vol. 26, n^{o} 1, p. 654-664, https://arxiv.org/abs/1907.12352 [*DOI :* 10.1109/TVCG.2019.2934334], https://hal.inria.fr/hal-02177779

[20] J. Jo, J. Seo, J.-D. Fekete.*PANENE: A Progressive Algorithm for Indexing and Querying Approximate k-Nearest Neighbors*, in "IEEE Transactions on Visualization and Computer Graphics", February 2020, vol. 26, n^{o} 2, p. 1347-1360 [*DOI :* 10.1109/TVCG.2018.2869149], https://hal.inria.fr/hal-01855672

[21] J. Jo, F. Vernier, P. Dragicevic, J.-D. Fekete.*A Declarative Rendering Model for Multiclass Density Maps*, in "IEEE Transactions on Visualization and Computer Graphics", January 2019, vol. 25, n^{o} 1, p. 470-480 [*DOI :* 10.1109/TVCG.2018.2865141], https://hal.inria.fr/hal-01848427

[22] M. Le Goc, C. Perin, S. Follmer, J.-D. Fekete, P. Dragicevic.*Dynamic Composite Data Physicalization Using Wheeled Micro-Robots*, in "IEEE Transactions on Visualization and Computer Graphics", January 2019, vol. 25, n^{o} 1, p. 737-747 [*DOI :* 10.1109/TVCG.2018.2865159], https://hal.inria.fr/hal-01848436

[23] D. Martín, G. Arroyo, V. del Sol, C. Romo, T. Isenberg.*Analysis of Drawing Characteristics for Reproducing Traditional Hand-Made Stippling*, in "Computers and Graphics", May 2019, vol. 80, p. 1–16 [*DOI :* 10.1016/j.cag.2019.02.001], https://hal.inria.fr/hal-02017607

[24] H. MIAO, T. KLEIN, D. KOUŘIL, P. MINDEK, K. SCHATZ, E. GRÖLLER, B. KOZLÍKOVÁ, T. ISENBERG, I. VIOLA.*Multiscale Molecular Visualization*, in "Journal of Molecular Biology", March 2019, vol. 431, n^{o} 6, p. 1049–1070 [*DOI :* 10.1016/J.JMB.2018.09.004], https://hal.inria.fr/hal-01873724

[25] N. TOVANICH, N. HEULOT, J.-D. FEKETE, P. ISENBERG.*Visualization of Blockchain Data: A Systematic Review*, in "IEEE Transactions on Visualization and Computer Graphics", 2020, forthcoming [*DOI :* 10.1109/TVCG.2019.2963018], https://hal.archives-ouvertes.fr/hal-02426339

[26] P. R. VALDIVIA, P. BUONO, C. PLAISANT, N. DUFOURNAUD, J.-D. FEKETE.*Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", 2020, forthcoming [*DOI :* 10.1109/TVCG.2019.2933196], https://hal.inria.fr/hal-02264960

[27] X. WANG, L. BESANÇON, M. AMMI, T. ISENBERG.*Augmenting Tactile 3D Data Navigation With Pressure Sensing*, in "Computer Graphics Forum", June 2019, vol. 38, n^{o} 3, p. 635-647 [*DOI :* 10.1111/CGF.13716], https://hal.archives-ouvertes.fr/hal-02091999

[28] Y. WANG, A. SEGAL, R. KLATZKY, D. F. KEEFE, P. ISENBERG, J. HURTIENNE, E. HORNECKER, T. DWYER, S. BARRASS.*An Emotional Response to the Value of Visualization*, in "IEEE Computer Graphics and Applications", September 2019, vol. 39, n^{o} 5, p. 8-17 [*DOI :* 10.1109/MCG.2019.2923483], https://hal.inria.fr/hal-02317718

Articles in Non Peer-Reviewed Journal

[29] J. ALEXANDER, P. ISENBERG, Y. JANSEN, B. E. ROGOWITZ, A. VANDE MOERE.*Data Physicalization*, in "Dagstuhl Reports", April 2019, vol. 8, n^{o} 10, p. 127-147 [*DOI :* 10.4230/DAGREP.8.10.127], https://hal.sorbonne-universite.fr/hal-02090877

[30] E. K. CHOE, R. DACHSELT, P. ISENBERG, B. LEE.*Mobile Data Visualization (Dagstuhl Seminar 19292)*, in "Dagstuhl Reports", 2019 [*DOI :* 10.4230/DAGREP.9.7.78], https://hal.inria.fr/hal-02434852

International Conferences with Proceedings

[31] L. BESANÇON, M. COOPER, A. YNNERMAN, F. VERNIER.*Population overlays on choropleths*, in "Journée Visu 2019", Paris, France, May 2019, https://hal.inria.fr/hal-02130444

[32] L. BESANÇON, P. DRAGICEVIC.*The Continued Prevalence of Dichotomous Inferences at CHI*, in "CHI '19 - Proceedings of CHI Conference on Human Factors in Computing Systems Extended Abstracts", Glasgow, United Kingdom, May 2019 [*DOI :* 10.1145/3290607.3310432], https://hal.inria.fr/hal-01980268

[33] T. BLASCHECK, B. SHARIF.*Visually Analyzing Eye Movements on Natural Language Texts and Source Code Snippets*, in "ETRA 2019 - ACM Symposium on Eye Tracking Research & Applications", Denver, United States, June 2019 [*DOI :* 10.1145/3314111.3319917], https://hal.inria.fr/hal-02084109

[34] *Best Paper*
P. DRAGICEVIC, Y. JANSEN, A. SARMA, M. KAY, F. CHEVALIER.*Increasing the Transparency of Research Papers with Explorable Multiverse Analyses*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019 [*DOI :* 10.1145/3290605.3300295], https://hal.inria.fr/hal-01976951.

[35] J.-D. FEKETE, Q. CHEN, Y. FENG, J. RENAULT.*Practical Use Cases for Progressive Visual Analytics*, in "DSIA 2019 - 4th Workshop on Data Systems for Interactive Analysis", Vancouver, Canada, October 2019, https://hal.inria.fr/hal-02342944

[36] J. FUCHS, P. ISENBERG, A. BEZERIANOS, M. MILLER, D. KEIM.*EduClust -A Visualization Application for Teaching Clustering Algorithms*, in "Eurographics - 40th Annual Conference of the European Association for Computer Graphics", Genova, Italy, May 2019, p. 1-8, Education paper, https://hal.inria.fr/hal-02103293

[37] U. OBAIDELLAH, M. RASCHKE, T. BLASCHECK.*Classification of Strategies for Solving Programming Problems using AoI Sequence Analysis*, in "ETRA 2019 - Symposium on Eye Tracking Research and Applications", Denver, United States, June 2019 [*DOI :* 10.1145/3314111.3319825], https://hal.inria.fr/hal-02084127

[38] N. SILVA, T. BLASCHECK, R. JIANU, N. RODRIGUES, D. WEISKOPF, M. RAUBAL, T. SCHRECK.*Eye Tracking Support for Visual Analytics Systems: Foundations, Current Applications, and Research Challenges*, in "ETRA 2019 - Symposium on Eye Tracking Research and Applications", Denver, United States, June 2019 [*DOI :* 10.1145/3314111.3319919], https://hal.inria.fr/hal-02084137

Conferences without Proceedings

[39] A. HOLCOMBE, C. LUDOWICI, S. HAROZ.*Is there a reproducibility crisis around here? Maybe not, but we still need to change*, in "Vision Sciences Society", St. Pete Beach, Florida, United States, 2019, https://hal.inria.fr/hal-01944513

[40] C. S. PETERSON, J. A. SADDLER, T. BLASCHECK, B. SHARIF.*Visually Analyzing Students' Gaze on C++ Code Snippets*, in "EMIP 2019 - 6th International Workshop on Eye Movements in Programming", Montreal, Canada, May 2019, https://hal.inria.fr/hal-02084148

[41] X. WANG, L. BESANÇON, F. GUÉNIAT, M. SERENO, M. AMMI, T. ISENBERG.*A Vision of Bringing Immersive Visualization to Scientific Workflows*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems - Workshop on Interaction Design & Prototyping for Immersive Analytics", Glasgow, United Kingdom, May 2019, https://hal.archives-ouvertes.fr/hal-02053969

Scientific Books (or Scientific Book chapters)

[42] P. DRAGICEVIC, Y. JANSEN, A. VANDE MOERE.*Data Physicalization*, in "Springer Handbook of Human Computer Interaction", J. VANDERDONCKT (editor), Springer Reference, Springer, December 2019, https://hal.inria.fr/hal-02113248

[43] J.-D. FEKETE, D. FISHER, A. NANDI, M. SEDLMAIR.*Progressive Data Analysis and Visualization*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, April 2019 [*DOI :* 10.4230/DAGREP.8.10.1], https://hal.inria.fr/hal-02090121

[44] J.-D. FEKETE, T. J. JANKUN-KELLY, M. TORY, K. XU.*Report from Dagstuhl Seminar 18462: Provenance and Logging for Sense Making*, Dagstuhl Publishing, April 2019, vol. 8, n° 11, p. 35–62 [*DOI :* 10.4230/DAGREP.8.11.35], https://hal.inria.fr/hal-02095768

[45] S. KNUDSEN, J. AERTS, D. ARCHAMBAULT, R. CHANG, J.-D. FEKETE, M. GRANDJEAN, J. KENNEDY, M. KIVELÄ, M. MAGNANI, H. PURCHASE, C. TOMINSKI, P. R. VALDIVIA, T. VON LANDESBERGER.*Unifying the framework of Multi-Layer Network and Visual Analytics*, in "Visual Analytics of Multilayer Networks Across Disciplines", Dagstuhl Reports, 2019, vol. 9, n° 2, p. 19-23, https://hal.archives-ouvertes.fr/hal-02185272

[46] B. E. ROGOWITZ, A. DIEHL, P. ISENBERG, R. BORGO, A. ABDUL-RAHMAN.*Marshalling the Many Facets of Diversity*, in "Diversity in Visualization", Synthesis Lectures on Visualization, Morgan Claypool, 2019 [*DOI :* 10.2200/S00894ED1V01Y201901VIS010], https://hal.inria.fr/hal-02317794

Research Reports

[47] S. HURON, N. BOUKHELIFA, T. BAUDEL, C. GOULARD, P. ISENBERG, J. TIERNY, R. VUILLEMOT.*Tendances de la visualisation en France*, UMR GMPA, AgroParisTech, INRA ; SID - Sociologie Information-Communication Design ; Télécom ParisTech ; SES - Département Sciences Economiques et Sociales ; Univ. Paris-Saclay, October 2019, https://hal.archives-ouvertes.fr/hal-02317029

Other Publications

[48] R. ARAVIND, T. BLASCHECK, P. ISENBERG.*A Survey on Sleep Visualizations for Fitness Trackers*, 2019, Posters of the European Conference on Visualization (EuroVis), Poster, https://hal.inria.fr/hal-02337783

[49] R. BLANCO, Z. SALAZAR, T. ISENBERG.*Exploring carnivorous plant habitats based on images from social media*, October 2019, IEEE VIS 2019: IEEE Conference on Visualization, Poster, https://hal.inria.fr/hal-02196764

[50] M. SERENO, L. BESANÇON, T. ISENBERG.*Supporting Volumetric Data Visualization and Analysis by Combining Augmented Reality Visuals with Multi-Touch Input*, June 2019, p. 21–23, EuroVis Posters, Poster [*DOI :* 10.2312/EURP.20191136], https://hal.inria.fr/hal-02123904

[51] *Best Paper*
N. TOVANICH, N. HEULOT, J.-D. FEKETE, P. ISENBERG.*A Systematic Review of Online Bitcoin Visualizations*, 2019, Posters of the European Conference on Visualization (EuroVis), The poster received the Honorable Mention award at EuroVis 2019 [*DOI :* 10.2312/EURP.20191148], https://hal.archives-ouvertes.fr/hal-02155171.

[52] C. TURKAY, N. PEZZOTTI, C. BINNIG, H. STROBELT, B. HAMMER, D. KEIM, J.-D. FEKETE, T. PALPANAS, Y. WANG, F. RUSU.*Progressive Data Science: Potential and Challenges*, July 2019, https://arxiv.org/abs/1812.08032 - working paper or preprint, https://hal.inria.fr/hal-01961871

References in notes

[53] S. K. CARD, J. D. MACKINLAY, B. SHNEIDERMAN (editors). *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999

[54] J. BERTIN.*Sémiologie graphique : Les diagrammes - Les réseaux - Les cartes*, Les réimpressions, Editions de l'Ecole des Hautes Etudes en Sciences, Paris, France, 1967

[55] A. BEZERIANOS, P. DRAGICEVIC, R. BALAKRISHNAN.*Mnemonic rendering: an image-based approach for exposing hidden changes in dynamic displays*, in "UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology", New York, NY, USA, ACM, 2006, p. 159–168, http://doi.acm.org/10.1145/1166253.1166279

[56] A. BEZERIANOS, P. DRAGICEVIC, J.-D. FEKETE, J. BAE, B. WATSON.*GeneaQuilts: A System for Exploring Large Genealogies*, in "IEEE Transactions on Visualization and Computer Graphics", October 2010, vol. 16, n^{o} 6, p. 1073-1081 [*DOI :* 10.1109/TVCG.2010.159], https://hal.inria.fr/inria-00532939

[57] J.-D. FEKETE, C. PLAISANT.*Interactive Information Visualization of a Million Items*, in "Proc. IEEE Symposium on Information Visualization 2002 (InfoVis 2002)", Boston, USA, IEEE Press, October 2002, p. 117-124

[58] J.-D. FEKETE, C. PLAISANT.*Les leçons tirées des deux compétitions de visualisation d'information*, in "Proceedings of IHM 2004", Namur, Belgium, International Conference Proceedings Series, ACM Press, September 2004, p. 7-12

[59] J. J. GIBSON.*The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates, New Jersey, USA, 1979

[60] Y. GUIARD, Y. DU, J.-D. FEKETE, M. BEAUDOUIN-LAFON, C. APPERT, O. CHAPUIS.*Shakespeare's Complete Works as a Benchmark for Evaluating Multiscale Document-Navigation Techniques*, in "Proceedings of BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV'06)", Venice, Italy, ACM Press, May 2006, p. 65-70

[61] N. HENRY RICHE, J.-D. FEKETE.*Evaluating Visual Table Data Understanding*, in "Proceedings of BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV'06)", Venice, Italy, ACM Press, May 2006, 6

[62] B. LEE, C. PLAISANT, C. SIMS PARR, J.-D. FEKETE, N. HENRY RICHE.*Task taxonomy for graph visualization*, in "BELIV '06: Proceedings of the 2006 AVI workshop on BEyond time and errors", New York, NY, USA, ACM, 2006, p. 1–5, http://doi.acm.org/10.1145/1168149.1168168

[63] C. PLAISANT, J.-D. FEKETE, G. GRINSTEIN.*Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository*, in "IEEE Transactions on Visualization and Computer Graphics", 2008, vol. 14, n^{o} 1, p. 120–134, http://doi.ieeecomputersociety.org/10.1109/TVCG.2007.70412

[64] V. SERRANO MOLINERO, B. BACH, C. PLAISANT, N. DUFOURNAUD, J.-D. FEKETE.*Understanding the Use of The Vistorian: Complementing Logs with Context Mini-Questionnaires*, in "Visualization for the Digital Humanities", Phoenix, United States, October 2017, https://hal.inria.fr/hal-01650259

[65] A. TRIESMAN.*Preattentive Processing in Vision*, in "Computer Vision, Graphics, and Image Processing", August 1985, vol. 31, n^{o} 2, p. 156-177

[66] E. TUFTE.*The Visual Display of Quantitative Information*, Graphics Press, 1983

[67] J. W. TUKEY.*Exploratory Data Analysis*, Addison-Wesley, 1977

Project-Team CEDAR

Rich Data Exploration at Cloud Scale

IN COLLABORATION WITH: Laboratoire d'informatique de l'école polytechnique (LIX)

IN PARTNERSHIP WITH:
Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Data and Knowledge Representation and Processing

Table of contents

Project-Team CEDAR

Creation of the Team: 2016 January 01, updated into Project-Team: 2018 April 01

Keywords:

Computer Science and Digital Science:

A3.1.1. - Modeling, representation
A3.1.2. - Data management, quering and storage
A3.1.3. - Distributed data
A3.1.6. - Query optimization
A3.1.7. - Open data
A3.1.8. - Big data (production, storage, transfer)
A3.1.9. - Database
A3.2.1. - Knowledge bases
A3.2.3. - Inference
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.3.1. - On-line analytical processing
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4.1. - Supervised learning
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A9.1. - Knowledge
A9.2. - Machine learning

Other Research Topics and Application Domains:

B8.5.1. - Participative democracy
B9.5.6. - Data science
B9.7.2. - Open data

1. Team, Visitors, External Collaborators

Research Scientists

Ioana Manolescu [Team leader, Inria, Senior Researcher, HDR]
Oana Balalau [Inria, Starting Research Position, from Nov 2019]
Yanlei Diao [École polytechnique, Researcher]

External Collaborators

Julien Leblay [AIST Japan, from May 2019]
Xavier Tannier [CNRS]

Technical Staff

Felipe Cordeiro Alves Dias [Inria, Engineer, from Apr 2019 until Aug 2019]
Laurent Cetinsoy [École polytechnique, Engineer, from May 2019 until Jul 2019]
Tayeb Merabti [Inria, Engineer]
Arnab Sinha [École polytechnique, Engineer, from Apr 2019]

PhD Students
Maxime Buron [Inria, PhD Student]
Tien Duc Cao [Inria, PhD Student, until Sep 2019]
Luciano Di Palma [École polytechnique, PhD Student]
Qi Fan [China Scholarship Council, PhD Student, from Dec 2019]
Paweł Guzewicz [École polytechnique, PhD Student]
Enhui Huang [École polytechnique, PhD Student]
Vincent Jacob [École polytechnique, PhD Student, from Dec 2019]
Felix Raimundo [École polytechnique, PhD Student, until Aug 2019]
Khaled Zaouk [École polytechnique, PhD Student]

Post-Doctoral Fellows
Mirjana Mazuran [Inria, Post-Doctoral Fellow]
Fei Song [École polytechnique, Post-Doctoral Fellow]

Visiting Scientist
Juliana Freire [Digiteo, until Jul 2019]

Administrative Assistant
Maeva Jeannot [Inria, Administrative Assistant, until Oct 2019]

2. Overall Objectives

2.1. Overall Objectives

Our research aims at **models, algorithms and tools for highly efficient, easy-to-use data and knowledge management**; throughout our research, **performance at scale** is a core concern, which we address, among other techniques, by designing algorithms for a **cloud (massively parallel)** setting. Our scientific contributions fall in three interconnected areas:

Expressive models for new applications As data and knowledge applications keep extending to novel application areas, we work to devise appropriate data and knowledge models, endowed with formal semantics, to capture such applications' needs. This work mostly concerns the domains of data journalism and journalistic fact checking;

Optimization and performance at scale This topic is at the heart of Y. Diao's ERC project "Big and Fast Data", which aims at optimization with performance guarantees for real-time data processing in the cloud. Machine learning techniques and multi-objectives optimization are leveraged to build performance models for data analytics the cloud. The same boal is shared by our work on efficient evaluation of queries in dynamic knowledge bases.

Data discovery and exploration Today's Big Data is complex; understanding and exploiting it is difficult. To help users, we explore: compact summaries of knowledge bases to abstrac their structure and help users formulate queries; interactive exploration of large relational databases; techniques for automatically discovering interesting information in knowledge bases; and keyword search techniques over Big Data sources.

3. Research Program

3.1. Scalable Heterogeneous Stores

Big Data applications increasingly involve *diverse* data sources, such as: structured or unstructured documents, data graphs, relational databases etc. and it is often impractical to load (consolidate) diverse data sources in a single repository. Instead, interesting data sources need to be exploited "as they are", with the added value of the data being realized especially through the ability to combine (join) together data from several sources.

Systems capable of exploiting diverse Big Data in this fashion are usually termed *polystores*. A current limitation of polystores is that data stays captive of its original storage system, which may limit the data exploitation performance. We work to devise highly efficient storage systems for heterogeneous data across a variety of data stores.

3.2. Semantic Query Answering

In the presence of data semantics, query evaluation techniques are insufficient as they only take into account the database, but do not provide the reasoning capabilities required in order to reflect the semantic knowledge. In contrast, (ontology-based) query answering takes into account both the data and the semantic knowledge in order to compute the full query answers, blending query evaluation and semantic reasoning.

We aim at designing efficient semantic query answering algorithms, both building on cost-based reformulation algorithms developed in the team and exploring new approaches mixing materialization and reformulation.

3.3. Multi-Model Querying

As the world's affairs get increasingly more digital, a large and varied set of data sources becomes available: they are either structured databases, such as government-gathered data (demographics, economics, taxes, elections, ...), legal records, stock quotes for specific companies, un-structured or semi-structured, including in particular graph data, sometimes endowed with semantics (see e.g. the Linked Open Data cloud). Modern data management applications, such as data journalism, are eager to combine in innovative ways both static and dynamic information coming from structured, semi-structured, and un-structured databases and social feeds. However, current content management tools for this task are not suited for the task, in particular when they require a lenghy rigid cycle of data integration and consolidation in a warehouse. Thus, we see a need for flexible tools allowing to interconnect various kinds of data sources and to query them together.

3.4. Interactive Data Exploration at Scale

In the Big Data era we are faced with an increasing gap between the fast growth of data and the limited human ability to comprehend data. Consequently, there has been a growing demand of data management tools that can bridge this gap and help users retrieve high-value content from data more effectively. To respond to such user information needs, we aim to build interactive data exploration as a new database service, using an approach called "explore-by-example".

3.5. Exploratory Querying of Semantic Graphs

Semantic graphs including data and knowledge are hard to apprehend for users, due to the complexity of their structure and oftentimes to their large volumes. To help tame this complexity, in prior research (2014), we have presented a full framework for RDF data warehousing, specifically designed for heterogeneous and semantic-rich graphs. However, this framework still leaves to the users the burden of chosing the most interesting warehousing queries to ask. More user-friendly data management tools are needed, which help the user discover the interesting structure and information hidden within RDF graphs. This research has benefitted from the arrival in the team of Mirjana Mazuran, as well as from the start of the PhD thesis of Paweł Guzewicz, co-advised by Yanlei Diao and Ioana Manolescu.

3.6. An Unified Framework for Optimizing Data Analytics

Data analytics in the cloud has become an integral part of enterprise businesses. Big data analytics systems, however, still lack the ability to take user performance goals and budgetary constraints for a task, collectively referred to as task objectives, and automatically configure an analytic job to achieve the objectives.

Our goal, is to come up with a data analytics optimizer that can automatically determine a cluster configuration with a suitable number of cores as well as other runtime system parameters that best meet the task objectives. To achieve this, we also need to design a multi-objective optimizer that constructs a Pareto optimal set of job configurations for task-specific objectives, and recommends new job configurations to best meet these objectives.

4. Application Domains

4.1. Cloud Computing

Cloud computing services are strongly developing and more and more companies and institutions resort to running their computations in the cloud, in order to avoid the hassle of running their own infrastructure. Today's cloud service providers guarantee machine availabilities in their Service Level Agreement (SLA), without any guarantees on performance measures according to a specific cost budget. Running analytics on big data systems require the user not to only reserve the suitable cloud instances over which the big data system will be running, but also setting many system parameters like the degree of parallelism and granularity of scheduling. Chosing values for these parameters, and chosing cloud instances need to meet user objectives regarding latency, throughput and cost measures, which is a complex task if it's done manually by the user. Hence, we need need to transform cloud service models from availabily to user performance objective rises and leads to the problem of multi-objective optimization. Research carried out in the team within the ERC project “Big and Fast Data Analytics” aims to develop a novel optimization framework for providing guarantees on the performance while controlling the cost of data processing in the cloud.

4.2. Computational Journalism

Modern journalism increasingly relies on content management technologies in order to represent, store, and query source data and media objects themselves. Writing news articles increasingly requires consulting several sources, interpreting their findings in context, and crossing links between related sources of information. CEDARresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. Within the ANR ContentCheck project, and also as part of our international collaboration with the AIST institute from Japan, we work on one hand, to lay down foundations for computational data journalism and fact checking, and also work to devise concrete algorithms and platforms to help journalists perform their work better and/or faster. This work is carried in collaboration with Le Monde's “Les Décodeurs”.

On a related topic, heterogeneous data integration under a virtual graph abstract model is studied within the ICODA Inria project which has started in September 2017. There, we collaborate with Les Décodeurs as well as with Ouest France and Agence France Presse (AFP). The data and knowledge integration framework resulting from this work will support journalists' effort to organize and analyze their knowledge and exploit it in order to produce new content.

5. Highlights of the Year

5.1. Highlights of the Year

- Through 2019 competitive **hiring**, the team has doubled its number of senior members: Oana Bălălău has been hired on an Inria Starting Researcher Position (SRP), and she joined in november; Angelos Anadiotis has been hired as a Gaspard Monge Assistant Professor at Ecole Polytechnique within the team.

- I. Manolescu and M. Buron have demonstrated the **ConnectionLens** system to the Defense Minister Florence Parly, as part of DataIA's showing for her visit at Inria, in April 2019 [0]. The national Inria director Bruno Sportisse, the military director of Ecole polytechnique François Bouchet, and the Fields medalist Cédric Villani were also present.
- As a member of the scientific committee of the **GFAIH** (Global Forum on AI for Humanity), I. Manolescu had the opportunity to meet, in a dinner at the Elysée Palace, and exchange with the French President Emmanuel Macron, the Economy and Industry Minister Bruno Le Maire, the Research Minister Frédérique Vidal, and the Digital Affairs Minister Cedric O [0].

5.1.1. Awards

- The demonstration "Spade: A Modular Framework for Analytical Exploration of RDF Graphs"[15] has obtained the **Best Demonstration Award** at the BDA conference 2019, where it has also been informally presented [0].

6. New Software and Platforms

6.1. Tatooine

KEYWORDS: RDF - JSon - Knowledge database - Databases - Data integration - Polystore

FUNCTIONAL DESCRIPTION: Tatooine allows to jointly query data sources of heterogeneous formats and data models (relations, RDF graphs, JSON documents etc.) under a single interface. It is capable of evaluating conjunctive queries over several such data sources, distributing computations between the underlying single-data model systems and a Java-based integration layer based on nested tuples.

- Participants: François Goasdoué, Ioana Manolescu, Javier Letelier Ruiz, Michaël Thomazo, Oscar Santiago Mendoza Rivera, Raphael Bonaque, Swen Ribeiro, Tien Duc Cao and Xavier Tannier
- Contact: Ioana Manolescu

6.2. AIDES

KEYWORDS: Data Exploration - Active Learning

FUNCTIONAL DESCRIPTION: AIDES is a data exploration software. It allows a user to explore a huge (tabular) dataset and discover tuples matching his or her interest. Our system repeatedly proposes the most informative tuples to the user, who must annotate them as "interesting" / "not-interesting", and as iterations progress an increasingly accurate model of the user's interest region is built. Our system also focuses on supporting low selectivity, high-dimensional interest regions.

- Contact: Yanlei Diao

6.3. OntoSQL

KEYWORDS: RDF - Semantic Web - Querying - Databases

[0] https://team.inria.fr/cedar/connectionlens/
[0] https://twitter.com/ioanamanol/status/1189478849651904513
[0] https://twitter.com/cedarinrialix/status/1185203276142256128

FUNCTIONAL DESCRIPTION: OntoSQL is a tool providing three main functionalities: - Loading RDF graphs (consisting of data triples and possibly a schema or ontology) into a relational database, - Saturating the data based on the ontology. Currently, RDF Schema ontologies are supported. - Querying the loaded data using conjunctive queries. Data can be loaded either from distinct files or from a single file containing them both. The loading process allows to choose between two storage schemas: - One triples table. - One table per role and concept. Querying provides an SQL translation for each conjunctive query according to the storage schema used in the loading process, then the SQL query is evaluated by the underlying relational database.

- Participants: Ioana Manolescu, Michaël Thomazo and Tayeb Merabti
- Partner: Université de Rennes 1
- Contact: Ioana Manolescu
- URL: https://ontosql.inria.fr/

6.4. ConnectionLens

KEYWORDS: Data management - Big data - Information extraction - Semantic Web

FUNCTIONAL DESCRIPTION: ConnectionLens treats a set of heterogeneous, independently authored data sources as a single virtual graph, whereas nodes represent fine-granularity data items (relational tuples, attributes, key-value pairs, RDF, JSON or XML nodes…) and edges correspond either to structural connections (e.g., a tuple is in a database, an attribute is in a tuple, a JSON node has a parent…) or to similarity (sameAs) links. To further enrich the content journalists work with, we also apply entity extraction which enables to detect the people, organizations etc. mentioned in text, whether full-text or text snippets found e.g. in RDF or XML. ConnectionLens is thus capable of finding and exploiting connections present across heterogeneous data sources without requiring the user to specify any join predicate.

- Contact: Manolescu Ioana
- Publication: ConnectionLens: Finding Connections Across Heterogeneous Data Sources
- URL: https://team.inria.fr/cedar/connectionlens/

6.5. INSEE-Extract

Spreadsheets extractor

KEYWORDS: RDF - Data extraction

FUNCTIONAL DESCRIPTION: Extract content of spreadsheets automatically and store it as RDF triples

- Participants: Ioana Manolescu, Xavier Tannier and Tien Duc Cao
- Contact: Tien Duc Cao
- Publication: Extracting Linked Data from statistic spreadsheets
- URL: https://gitlab.inria.fr/cedar/excel-extractor

6.6. INSEE-Search

KEYWORDS: Document ranking - RDF

FUNCTIONAL DESCRIPTION: Searching for relevant data cells (or data row/column) given a query in natural language (French)

- Participants: Ioana Manolescu, Xavier Tannier and Tien Duc Cao
- Contact: Tien Duc Cao
- Publications: Extracting Linked Data from statistic spreadsheets - Searching for Truth in a Database of Statistics

6.7. RDFQuotient

Quotient summaries of RDF graphs

KEYWORDS: RDF - Graph algorithmics - Graph visualization - Graph summaries - Semantic Web

FUNCTIONAL DESCRIPTION: RDF graphs can be large and heterogeneous, making it hard for users to get acquainted with a new graph and understand whether it may have interesting information. To help users figure it out, we have devised novel equivalence relations among RDF nodes, capable of recognizing them as equivalent (and thus, summarize them together) despite the heterogeneity often exhibited by their incoming and outgoing node properties. From these relations, we derive four novel summaries, called Weak, Strong, Typed Weak and Typed Strong, and show how to obtain from them compact and enticing visualizations.

- Participants: Ioana Manolescu, Pawel Guzewicz and François Goasdoué
- Partner: Université de Rennes 1
- Contact: Manolescu Ioana
- Publications: hal-01325900v6 - Structural Summarization of Semantic Graphs

6.8. AIDEme

KEYWORDS: Active Learning - Data Exploration

SCIENTIFIC DESCRIPTION: AIDEme is a large-scale interactive data exploration system that is cast in a principled active learning (AL) framework: in this context, we consider the data content as a large set of records in a data source, and the user is interested in some of them but not all. In the data exploration process, the system allows the user to label a record as “interesting” or “not interesting” in each iteration, so that it can construct an increasingly-more-accurate model of the user interest. Active learning techniques are employed to select a new record from the unlabeled data source in each iteration for the user to label next in order to improve the model accuracy. Upon convergence, the model is run through the entire data source to retrieve all relevant records.

A challenge in building such a system is that existing active learning techniques experience slow convergence in learning the user interest when such exploration is performed on large datasets: for example, hundreds of labeled examples are needed to learn a user interest model over 6 attributes, as we showed using a digital sky survey of 1.9 million records. AIDEme employs a set of novel techniques to overcome the slow convergence problem:

• Factorization: We observe that a user labels a data record, her decision making process often can be broken into a set of smaller questions, and the answers to these questions can be combined to derive the final answer. This insight, formally modeled as a factorization structure, allows us to design new active learning algorithms, e.g., factorized version space algorithms [2], that break the learning problem into subproblems in a set of subspaces and perform active learning in each subspace, thereby significantly expediting convergence.

• Optimization based on class distribution: Another interesting observation is that when projecting the data space for exploration onto a subset of dimensions, the user interest pattern projected onto such a subspace often entails a convex object. When such a subspatial convex property holds, we introduce a new “dual-space model” (DSM) that builds not only a classification model from labeled examples, but also a polytope model of the data space that offers a more direct description of the areas known to be positive, areas known to be negative, and areas with unknown labels. We use both the classification model and the polytope model to predict unlabeled examples and choose the best example to label next. • Formal results on convergence: We further provide theoretical results on the convergence of our proposed techniques. Some of them can be used to detect convergence and terminate the exploration process. • Scaling to large datasets: In many applications the dataset may be too large to fit in memory. In this case, we introduce subsampling procedures and provide provable results that guarantee the performance of the model learned from the sample over the entire data source.

FUNCTIONAL DESCRIPTION: There is an increasing gap between fast growth of data and limited human ability to comprehend data. Consequently, there has been a growing demand for analytics tools that can bridge this gap and help the user retrieve high-value content from data. We introduce AIDEme, a scalable interactive data exploration system for efficiently learning a user interest pattern over a large dataset. The system is cast in a principled active learning (AL) framework, which iteratively presents strategically selected records for user labeling, thereby building an increasingly-more-accurate model of the user interest. However, a challenge in building such a system is that existing active learning techniques experience slow convergence when learning the user interest on large datasets. To overcome the problem, AIDEme explores properties of the user labeling process and the class distribution of observed data to design new active learning algorithms, which come with provable results on model accuracy, convergence, and approximation, and have evaluation results showing much improved convergence over existing AL methods while maintaining interactive speed.

RELEASE FUNCTIONAL DESCRIPTION: Project code can be found over: https://gitlab.inria.fr/ldipalma/aideme

- Participants: Luciano Di Palma and Enhui Huang
- Contact: Yanlei Diao
- URL: http://www.lix.polytechnique.fr/aideme

7. New Results

7.1. Quotient summaries of RDF graphs

We have continued and finalized our work on the question of efficiently computing informative summaries of large, heterogeneous RDF graphs. Such summaries simplify the users' efforts to understand and grasp the content of an RDF graph with which they are not familiar. For instance, Figure 1 shows the summary constructed fully automatically out of a benchmark graph of a bit more than 100 million triples.

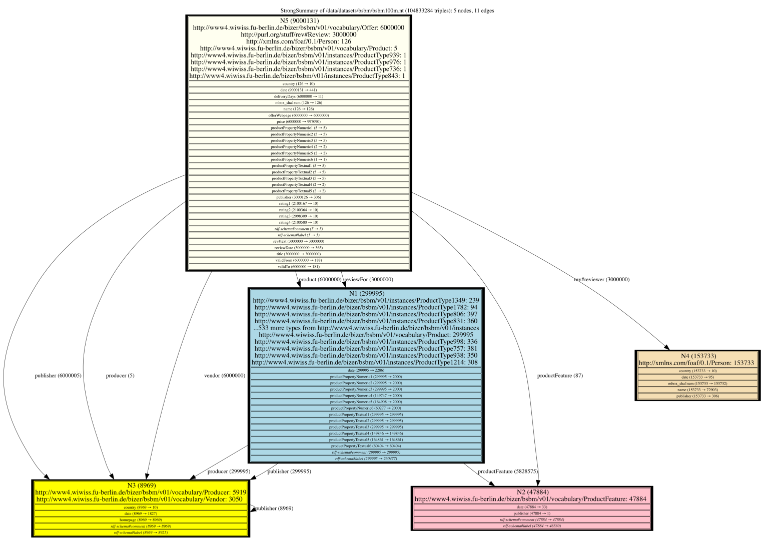


Figure 1. RDFQuotient summary of a 100 million triples graph.

We have presented, together with co-authors, a tutorial on the problem of summarizing RDF graphs, at the EDBT 2019 conference [21].

We have demonstrated new algorithms for efficiently building RDF quotient summaries out of large RDF graphs, in an incremental fashion, in [19].

Last but not least, a VLDB Journal submitted article systematizing most of our contributions in this area has been accepted (pending a minor, strictly cosmetic revision which will be sent out in January 2020).

7.2. Efficient query answering over semantic graphs

Query answering in RDF knowledge bases has traditionally been performed either through graph saturation, that is, adding all implicit triples to the graph, or through query reformulation, i.e. modifying the query to look for the explicit triples entailing precisely what the original query asks for. The most expressive fragment of RDF for which reformulation-based quey answering exists is the so-called database fragment of RDF (Goasdoué et al., EDBT 2013), in which implicit triples are restricted to those entailed using an RDFS ontology. Within this fragment, query answering was so far limited to the interrogation of data triples (non-RDFS ones); however, a powerful feature specific to RDF is the ability to query data and schema triples together. In [12], we address the general query answering problem by reducing it, through a pre-query reformulation step, to that solved by the query reformulation technique mentioned above (EDBR 2013). Our experiments also demonstrate the very modest cost (performance overhead) of this more powerful (more expressive) reformulation algorithm.

7.3. Scalable storage for polystores

Big data applications routinely involve diverse datasets: relations flat or nested, complex-structure graphs, documents, poorly structured logs, or even text data. To handle the data, application designers usually rely on several data stores used side-by-side, each capable of handling one or a few data models (e.g., many relational stores can also handle JSON data), and each very efficient for some, but not all, kinds of processing on the data.

A current limitation is that applications are written taking into account which part of the data is stored in which store and how. This fails to take advantage of (i) possible redundancy, when the same data may be accessible (with different performance) from distinct data stores; (ii) partial query results (in the style of materialized views) which may be available in the stores. If data migrates to another store, to take advantage of its performance for a specific task, applications must be re-written; this is tedious and error-prone.

In [11], we present ESTOCADA, a novel approach connecting applications to the potentially heterogeneous systems where their input data resides. ESTOCADA can be used in a polystore setting to transparently enable each query to benefit from the best combination of stored data and available processing capabilities. ESTOCADA leverages recent advances in the area of view-based query rewriting under constraints, which we use to describe the various data models and stored data. Our experiments illustrate the significant performance gains achieved by ESTOCADA.

7.4. Novel fact-checking architectures and algorithms

A frequent journalistic fact-checking scenario is concerned with the **analysis of statements** made by individuals, whether in public or in private contexts, and the propagation of information and hearsay ("who said/knew what when"), mostly in the public sphere (e.g., in discourses, statements to the media, or on public social networks such as Twitter), but also in private contexts (these become accessible to journalists through their sources). Inspired by our collaboration with fact-checking journalists from Le Monde, France's leading newspaper, we have described in [17] a Linked Data (RDF) model, endowed with formal foundations and semantics, for describing *facts, statements*, and *beliefs*. Our model combines temporal and belief dimensions to trace propagation of knowledge between agents along time, and can answer a large variety of interesting questions through RDF query evaluation. A preliminary feasibility study of our model incarnated in a corpus of tweets demonstrates its practical interest.

Based on the above model, we implemented and demonstrated BELINK [13], a prototype capable of storing such interconnected corpora, and answer powerful queries over them relying on SPARQL 1.1. The demo showcased the exploration of a rich real-data corpus built from Twitter and mainstream media, and interconnected through extraction of statements with their sources, time, and topics.

Statistic (numerical) data, e.g., on unemployment rates or immigrant populations, are hot fact-checking topics. In prior work, we have transformed a corpus of high-quality statistics from INSEE, the French national statistics institute, into an RDF dataset (Cao et al., Semantic Big Data Workshop, 2017, https://hal.inria.fr/hal-01583975), and shown how to locate inside the information most relevant to (thus, most likely to be useful to fact-check) a given keyword query (Cao et al., Web and Databases Workshop, 2018, https://hal.inria.fr/hal-01745768). Following on the above work, in [16], we present a novel approach to extract from text documents, e.g., online media articles, mentions of statistic entities from a reference source. A claim states that an entity has certain value, at a certain time. This completes a fact-checking pipeline from text, to the reference data closest to the claim. Using it, fact-checking journalists only have to interpret the difference between the claimed and the reference value. We evaluated our method on the INSEE reference dataset and show that it is efficient and effective. Further, this algorithm was adapted also to the (more challenging) context of content published on Twitter. This has lead to a semi-automatic interface for detecting statistic claims made in tweets and starting a semi-automatic fact-check of those claims, based on INSEE data. Figure 2 depicts the interface of this Twitter fact-checking system, which was shared with our Le Monde journalist partners.

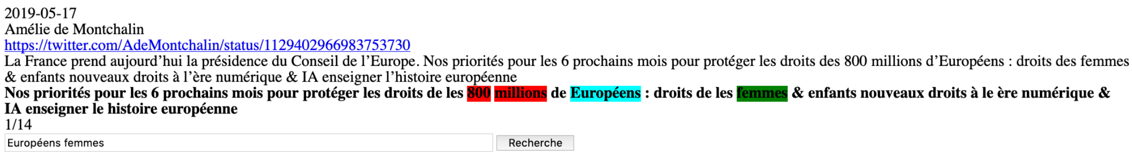


Figure 2. Screen capture of our Twitter fact-checking module.

7.5. Semantic graph exploration through interesting aggregates

RDF graphs can be large and complex; finding out interesting information within them is challenging. One easy method for users to discover such graphs is to be shown *interesting aggregates* (under the form of two-dimensional graphs, i.e., bar charts), where interestingness is evaluated through statistics criteria. While well understood for relational data, such exploration raises multiple challenges for RDF: facts, dimensions and measures have to be *identified* (as opposed to known beforehand); as there are more candidate aggregates, assessing their interestingness can be very costly; finally, *ontologies* bring novel specific challenges through the presence of *implicit* data, but also novel opportunities, enabling *ontology-driven exploration* from an aggregate initially proposed by the system.

The system DAGGER we had previously proposed (2017) pioneered this approach, however its is quite inefficient, in particular due to the need to evaluate numerous, expensive aggregation queries.

In 2019, we have built upon DAGGER to develop more efficient and more expressive versions thereof. Thus:

- In [22], we describe DAGGER^{+}, which builds upon DAGGER and leverages *sampling* to speed up the evaluation of potentially interesting agregates. We show that DAGGER^{+} achieves very significant execution time reductions, while reaching results very close to those of the original, less efficient system.
- Going beyond the expressive power of (candidate aggregates enumerated by) DAGGER, we have developed and demonstrated [15] SPADE, a *generic, extensible framework*, which we instantiated with:

(i) novel methods for enumerating candidate measures and dimensions in the vast space of possibilities provided by an RDF graph; (ii) a set of aggregate interestingness functions; (iii) ontology-based interactive exploration, and (iv) efficient early-stop techniques for estimating the interestingness of an aggregate query. A multi-dimensional aggregate automatically identified by SPADE appears in Figure 3.

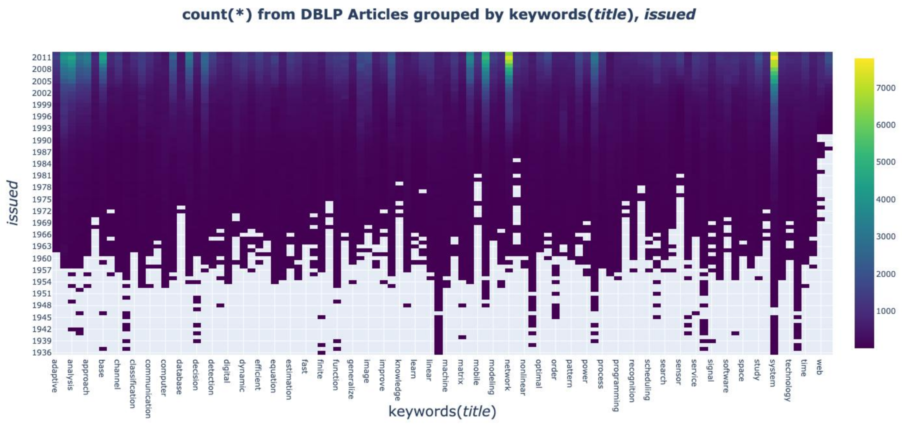


Figure 3. Interesting multi-dimensional aggregate automatically identified by DAGGER.

7.6. A Next-Generation Unified Data Analytics Optimizer

Big data analytics systems today still lack the ability to take user performance goals and budgetary constraints, collectively referred to as "objectives", and automatically configure an analytic job to achieve the objectives.

In [10], we present a unified data analytics optimizer that can automatically determine the parameters of the runtime system, collectively called a job configuration, for general dataflow programs based on user objectives. UDAO embodies key techniques including in-situ modeling, which learns a model for each user objective in the same computing environment as the job is run, and multi-objective optimization, which computes a Pareto optimal set of job configurations to reveal tradeoffs between different objectives.

Using benchmarks developed based on industry needs, our demonstration will allow the user to explore (1) learned models to gain insights into how various parameters affect user objectives; (2) Pareto frontiers to understand interesting tradeoffs between different objectives and how a configuration recommended by the optimizer explores these tradeoffs; (3) end- to-end benefits that UDAO can provide over default configurations or those manually tuned by engineers.

We demonstrated this work at the VLDB 2019 conference.

7.7. A factorized version space algorithm for interactive database exploration

One challenge in building an interactive database exploration system is that existing active learning (AL) techniques experience slow convergence when learning the user interest on large datasets. To address this slow convergence problem, we augmented version space-based AL algorithms, which have strong theoretical results on convergence but are very costly to run, with additional insights obtained in the user labeling process. These insights lead to a novel algorithm that factorizes the version space to perform active learning in a set of subspaces, with provable results on optimality, as well as optimizations for better performance. Evaluation results using real world datasets show that our algorithm significantly outperforms state-of-the-art version space algorithms, as well as our previous data exploration algorithm DSM (Huang et al., PVLDB 2018), for large database exploration.

The above work was accepted as a conference paper at ICDM 2019 [14]. In addition, we have presented a demonstration of our software at NeurIPS 2019 [26], where people could interact with our system over two real-world datasets, and also observe how our system compares against traditional AL algorithms.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

- AIDE ("A New Database Service for Interactive Exploration on Big Data") is an ANR "Young Researcher" project led by Y. Diao, started at the end of 2016.
- ContentCheck (2015-2018) is an ANR project led by I. Manolescu, in collaboration with U. Rennes 1 (F. Goasdoué), INSA Lyon (P. Lamarre), the LIMSI lab from U. Paris Sud, and the Le Monde newspaper, in particular their fact-checking team Les Décodeurs. Its aim is to investigate content management models and tools for journalistic fact-checking.
- CQFD (2019-2022) is an ANR project coordinated by F. Ulliana (U. Montpellier), in collaboration with U. Rennes 1 (F. Goasdoué), Inria Lille (P. Bourhis), Institut Mines Télécom (A. Amarilli), Inria Paris (M. Thomazo) and CNRS (M. Bienvenu). Its research aims at investigating efficient data management methods for ontology-based access to heterogeneous databases (polystores).

8.1.2. Others

- The goal of the iCODA project is to develop the scientific and technological foundations for knowledge- mediated user-in-the-loop collaborative data analytics on heterogenous information sources, and to demonstrate the effectiveness of the approach in realistic, high-visibility use-cases. The project stands at the crossroad of multiple research fields—content analysis, data management, knowledge represen- tation, visualization—that span multiple Inria themes, and counts on a club of major press partners to define usage scenarios, provide data and demonstrate achievements. This is a project funded directly by Inria ("Inria Project Lab"), and is in collaboration with GraphIK, ILDA, LINKMEDIA (coordinator), as well as the press partners AFP, Le Monde (Les Décodeurs) and Ouest-France.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

IDEAA: Issue-Driven European Arena Analytics is a project funded by the European Commission Union's Horizon 2020 research and innovation programme. The project started in July 2018 for a duration of two years. Its purpose is to allow citizens to easily explore the trove of publicly available data with the aim of building a viewpoint on specific issues. Its main strengths are: supply users with succinct and meaningful knowledge with respect to the issue they are interested in; allow users to interact with the provided knowledge to refine their information need and advance understanding; suggest interesting or unexpected aspects in the data and support the comparison of knowledge discovered from different data sources. IDEAA is inspired by human-to-human dialogues, where questions are explorative, possibly imprecise, and answers may be a bit inaccurate but suggestive, conveying an idea that stimulates the interlocutor to further questions.

The project supports a two-years presence of Mirjana Mazuran as an experienced post-doc in our team.

8.3. International Initiatives

8.3.1. Inria Associate Teams Not Involved in an Inria International Labs

8.3.1.1. WebClaimExplain

Title: Mining for explanations to claims published on the Web

International Partner (Institution - Laboratory - Researcher):

AIST (Japan) - Julien Leblay

Start year: 2017

See also: https://team.inria.fr/cedar/projects/webclaimexplain/

The goal of this research is to create tools to find explanations for facts and verify claims made online. While this process cannot be fully automated, the main focus of our work will be explanation finding via trusted sources, based on the observation that one can only trust a statement if he/she can explain it through rules and proofs that can themselves be trusted.

8.3.2. Inria International Partners

8.3.2.1. Informal International Partners

- We collaborate with Alin Deutsch and Rana Al-Otaibi from the University of California in San Diego, on the topic of efficient data management in polystore sytems.
- We collaborate with Helena Galhardas from the University of Lisbon on the topic of efficiently interconnecting heterogeneous data sources for journalistic applications.
- We collaborate with Anna Liu from U. Massachussets at Amherst; she co-advises PhD thesis of several students in the group (E. Huang and L. Di Palma).

8.3.3. Participation in International Programs

8.3.3.1. AYAME

WebClaimExplain

Title: Mining for explanations to claims published on the Web

International Partner (Institution - Laboratory - Researcher):

AIST (Japan) - Leblay Julien

Duration: 2017 - 2019

Start year: 2017

See also: https://team.inria.fr/cedar/connectionlens/

The goal of this research is to create tools to find explanations for facts and verify claims made online. While this process cannot be fully automated, the main focus of our work will be explanation finding via trusted sources, based on the observation that one can only trust a statement if he/she can explain it through rules and proofs that can themselves be trusted.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

We have hosted from January to July 2019 the sabbatical visit of Juliana Freire, a professor at the New York University and the president of the prestigious ACM SIGMOD scientific association.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

I. Manolescu was a steering committee member for the International Workshop on Misinformation, Computational Fact-Checking and Credible Web in conjunction with The Web Conference 2019.

I. Manolescu was a member of the scientific committee in charge of organizing the Global Forum on AI for Humanity (http://gfaih.org), an international conference organized under the patronage of the French government to discuss the impact and perspectives for AI research on science and the society at large. The conference featured an opening intervention by Cédric Villani and a closing speech by the French president Emmanuel Macron.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

I. Manolescu has been a chair of the tutorial track at the ACM SIGMOD 2019 conference.

9.1.2.2. Member of the Conference Program Committees

I. Manolescu has been a member of the program committees of: the IEEE International Conference on Data Engineering (ICDE, demonstrations track) 2019, the DASFAA Conference 2019, the Extended Semantic Web Conference (ESWC) 2019, and the International Conference on Web Engineering (ICWE) 2019.

9.1.3. Journal

Y. Diao has been the Editor-in-Chief of the ACM SIGMOD Record.

9.1.3.1. Member of the Editorial Boards

Y. Diao has been an associate editor of the ACM Transactions on Database Systems (TODS).

I. Manolescu has been a member of the editorial board of the Proceedings of Very Large Databases (PVLDB) journal.

9.1.4. Invited Talks

I. Manolescu has given the following **keynote talks**:

- "Journalistic Dataspaces: Data Management for Journalism and Fact-Checking", keynote talk at the EDBT (Extending Database Technologies) Conference 2019 [28].
- "Computational fact-checking: problems, state of the art and perspectives", keynote at EGC (Extraction and Gestion de Connaissances, the French-speaking knowledge extraction and knowledge management conference) 2019 [27].

9.1.5. Leadership within the Scientific Community

- Y. Diao and I. Manolescu are members of the **PVLDB Endowment Board**, the entity in charge of organizing the publication of the prestigious PVLDB journal (A* in the CORE ranking) and of organizing the yearly PVLDB conference.
- Y. Diao has been the Chair of the ACM SIGMOD Research Highlight Award, a member of the ACM SIGMOD Executive Committee, and a member of the ACM SIGMOD Software Systems Award Committee.
- I. Manolescu is a member of the steering committee of **BDA**, the entity in charge of organizing: the yearly informal Bases de Données Avancées (BDA) conference, mostly attended by members of the French-speaking data management scientific community; and a summer school on Big Data Management, every two years.

9.1.6. Scientific Expertise

I. Manolescu has been part of the HCERES visiting committee of the Laboratoire Informatique de Grenoble (LIG) on December 2-4.

9.1.7. Research Administration

I. Manolescu has become the scientific director of **LabIA**, an initiative by the DINUM (Direction Interministerielle du Numérique) whose goal is to apply AI research and technology solutions to problems raised by the public administration, at the local or regional level. LabIA ran a selective application process which funded a dozen projects to be carried over by technology company (contractors) and four to be solved by research teams working together with the promoters (teams involved in public administration). The research projects funded by LabIA are respectively proposed by: the Cour de Cassation (the highest jurisdiction of the state), the Direction Générale de Controle de la Concurrence et de la Repression des Fraudes (DGCCRF, the national consumer watchdog agency), la SHOM (Service Hydrographique de la Marine, the seabed mapping service of the Marine) and the IGN (Institut Géographique National), in particular the team that is in charge of producing the detailed, dynamic information of the positioning of every fragment in the Earth crust.

I. Manolescu has been a member of **Inria Commission d'Evaluation** until the summer of 2019. As a consequence, she participated to the hiring committees for junior researchers (CRCN) of the Inria Lille and Inria Grenoble research centers, in May 2019; she has also participated to the final executive committee meeting that decided on the hires, in Paris, in June 2019.

I. Manolescu has been a member of a hiring committee that recruited a full-time Assistant Professor in Data Management at **Ecole Polytechnique**, and she has also headed another committee that recruited a part-time Assistant Professor in Data Science at Ecole Polytechnique.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

I. Manolescu is a part-time (50%) professor at Ecole Polytechnique, where she teaches:

- Master: I. Manolescu, "Database Management Systems", 52h, M1, École Polytechnique.
- Licence: I. Manolescu, "Giant Global Graph", 18h, L3, École Polytechnique.

She also teaches on appointment outside of Ecole Polytechnique:

- Master: I. Manolescu, "Architectures for Massive Data Management", 20h, M2, Université Paris-Saclay.

M. Buron and P. Guzewicz are Teaching Assistants at Ecole Polytechnique. Further, P. Guzewicz also taught 12h of lab in the M2 course "Architectures for Massive Data Management" mentioned above.

9.2.2. Supervision

PhD in progress: Maxime Buron: "Raisonnement efficace sur des grands graphes hétérogènes", since October 2017, François Goasdoué, Ioana Manolescu and Marie-Laure Mugnier (GraphIK Inria team in Montpellier)

PhD: Tien-Duc Cao, "Toward Automatic Fact-Checking of Statistic Claims", Université de Paris Saclay, 26/09/2019, Ioana Manolescu and Xavier Tannier (LIMICS, Université de Paris-Sorbonne).

PhD in progress: Ludivine Duroyon: "Data management models, algorithms & tools for fact-checking", since October 2017, François Goasdoué and Ioana Manolescu (Ludivine is in the Shaman team of U. Rennes 1 and IRISA, in Lannion)

PhD in progress: Paweł Guzewicz: "Expressive and efficient analytics for RDF graphs", since October 2018, Yanlei Diao and Ioana Manolescu.

PhD in progress: Qi Fan: "Multi-Objective Optimization for Data Analytics in the Cloud", since December 2019, Yanlei Diao.

PhD in progress: Enhui Huang: "Interactive Data Exploration at Scale", since October 2016, Yanlei Diao and Anna Liu (U. Massachussets at Amherst, USA).

PhD in progress: Vincent Jacob: "Explainable Anomaly Detection in High-Volume Stream Analytics", since December 2019, Yanlei Diao.

PhD in progress: Luciano di Palma, “New sampling algorithms and optimizations for interactive exploration in Big Data", since October 2017, Yanlei Diao and Anna Liu (U. Massachussets at Amherst, USA)

PhD in progress: Khaled Zaouk: “Performance Modeling and Multi-Objective Optimization for Data Analytics in the Cloud", since October 2017, Yanlei Diao.

9.2.3. Juries

- I. Manolescu has been part of the PhD committee of Adnène Belfodil, who defended his PhD thesis titled “Exceptional Model Mining for Behavioral Data Analysis” at INSA Lyon, on October 24, 2019.

9.3. Popularization

9.3.1. Articles and contents

I. Manolescu has been interviewed in the following general-audience media publications:

- “L’intelligence artificielle signe-t-elle la fin du journalisme ?”, Science et et Avenir special issue on IA, Sept 25 (dated November) 2019
- “Fake news: ces technologies qui les traquent”, Industrie et Technologies, Feb 5, 2019
- “Les algorithmes à l’assaut de la désinformation”, Science et Avenir, January 29, 2019
- “Les seniors partagent sept fois plus de «fake news» que les jeunes sur Facebook”, in Le Figaro, January 2019

9.3.2. Interventions

- Ioana Manolescu participated to a social science conference "Post-vérité et intox: où allons-nous?", organized by Fondation Maison des sciences de l’homme (FMSH) and Cité des Sciences, in February 2019 (presentation slides, présentation video)
- I. Manolescu presented her career and research at the “*Rendez-vous des jeunes mathématiciennes et informaticiennes*” (RJMI, a math and CS event organized for high-school female students) in October 2019.
- M. Buron, V. Jacob and I. Manolescu presented data management research to a group of 6 interns (13 years old, one-week long *stage de 3e*) in December 2019.

10. Bibliography

Major publications by the team in recent years

[1] R. Alotaibi, D. Bursztyn, A. Deutsch, I. Manolescu, S. Zampetakis.*Towards Scalable Hybrid Stores: Constraint-Based Rewriting to the Rescue*, in "SIGMOD 2019 - ACM SIGMOD International Conference on Management of Data", Amsterdam, Netherlands, June 2019, https://hal.inria.fr/hal-02070827

[2] M. Buron, F. Goasdoué, I. Manolescu, M.-L. Mugnier.*Reformulation-based query answering for RDF graphs with RDFS ontologies*, in "ESWC 2019 - European Semantic Web Conference", Portoroz, Slovenia, March 2019, https://hal.archives-ouvertes.fr/hal-02051413

[3] D. Bursztyn, F. Goasdoué, I. Manolescu.*Teaching an RDBMS about ontological constraints*, in "Very Large Data Bases", New Delhi, India, September 2016, https://hal.inria.fr/hal-01354592

[4] S. CAZALENS, P. LAMARRE, J. LEBLAY, I. MANOLESCU, X. TANNIER.*A Content Management Perspective on Fact-Checking*, in "The Web Conference 2018 - alternate paper tracks "Journalism, Misinformation and Fact Checking"", Lyon, France, April 2018, p. 565-574, https://hal.archives-ouvertes.fr/hal-01722666

[5] S. CEBIRIC, F. GOASDOUÉ, H. KONDYLAKIS, D. KOTZINOS, I. MANOLESCU, G. TROULLINOU, M. ZNEIKA.*Summarizing Semantic Graphs: A Survey*, in "The VLDB Journal", 2018, forthcoming, https://hal.inria.fr/hal-01925496

[6] Y. DIAO, P. GUZEWICZ, I. MANOLESCU, M. MAZURAN.*Spade: A Modular Framework for Analytical Exploration of RDF Graphs*, in "VLDB 2019 - 45th International Conference on Very Large Data Bases", Los Angeles, United States, Proceedings of the VLDB Endowment, Vol. 12, No. 12, August 2019 [*DOI :* 10.14778/3352063.3352101], https://hal.inria.fr/hal-02152844

[7] E. HUANG, L. PENG, L. D. PALMA, A. ABDELKAFI, A. LIU, Y. DIAO.*Optimization for active learning-based interactive database exploration*, in "Proceedings of the VLDB Endowment (PVLDB)", September 2018, vol. 12, n^{o} 1, p. 71-84 [*DOI :* 10.14778/3275536.3275542], https://hal.inria.fr/hal-01969886

[8] A. ROY, Y. DIAO, U. EVANI, A. ABHYANKAR, C. HOWARTH, R. LE PRIOL, T. BLOOM.*Massively Parallel Processing of Whole Genome Sequence Data: An In-Depth Performance Study*, in "SIGMOD '17 Proceedings of the 2017 ACM International Conference on Management of Dat", Chicago, Illinois, United States, SIGMOD '17 Proceedings of the 2017 ACM International Conference on Management of Data, ACM, May 2017, p. 187-202 [*DOI :* 10.1145/3035918.3064048], https://hal.inria.fr/hal-01683398

Publications of the year

Articles in International Peer-Reviewed Journal

[9] S. CEBIRIC, F. GOASDOUÉ, H. KONDYLAKIS, D. KOTZINOS, I. MANOLESCU, G. TROULLINOU, M. ZNEIKA.*Summarizing Semantic Graphs: A Survey*, in "The VLDB Journal", June 2019, vol. 28, n^{o} 3, https://hal.inria.fr/hal-01925496

[10] K. ZAOUK, F. SONG, C. LYU, A. SINHA, Y. DIAO, P. SHENOY.*UDAO: A Next-Generation Unified Data Analytics Optimizer*, in "Proceedings of the VLDB Endowment (PVLDB)", August 2019, vol. 12, n^{o} 12 [*DOI :* 10.14778/3352063.3352103], https://hal.inria.fr/hal-02267180

International Conferences with Proceedings

[11] R. ALOTAIBI, D. BURSZTYN, A. DEUTSCH, I. MANOLESCU, S. ZAMPETAKIS.*Towards Scalable Hybrid Stores: Constraint-Based Rewriting to the Rescue*, in "SIGMOD 2019 - ACM SIGMOD International Conference on Management of Data", Amsterdam, Netherlands, June 2019, https://hal.inria.fr/hal-02070827

[12] M. BURON, F. GOASDOUÉ, I. MANOLESCU, M.-L. MUGNIER.*Reformulation-based query answering for RDF graphs with RDFS ontologies*, in "ESWC 2019 - European Semantic Web Conference", Portoroz, Slovenia, March 2019, https://hal.archives-ouvertes.fr/hal-02051413

[13] T.-D. CAO, L. DUROYON, F. GOASDOUÉ, I. MANOLESCU, X. TANNIER.*BeLink: Querying Networks of Facts, Statements and Beliefs*, in "ACM CIKM : 28th International Conference on Information and Knowledge Management", Beijing, China, November 2019 [*DOI :* 10.1145/3357384.3357851], https://hal.inria.fr/hal-02269134

[14] L. Di Palma, Y. Diao, A. Liu.*A Factorized Version Space Algorithm for "Human-In-the-Loop" Data Exploration*, in "ICDM - 19th IEEE International Conference in Data Mining", Beijing, China, November 2019, https://hal.inria.fr/hal-02274497

[15] Y. Diao, P. Guzewicz, I. Manolescu, M. Mazuran.*Spade: A Modular Framework for Analytical Exploration of RDF Graphs*, in "VLDB 2019 - 45th International Conference on Very Large Data Bases", Los Angeles, United States, Proceedings of the VLDB Endowment, Vol. 12, No. 12, August 2019 [*DOI :* 10.14778/3352063.3352101], https://hal.inria.fr/hal-02152844

[16] T. Duc Cao, I. Manolescu, X. Tannier.*Extracting statistical mentions from textual claims to provide trusted content*, in "NLDB 2019 - 24th International Conference on Applications of Natural Language to Information Systems", Salford, United Kingdom, June 2019, https://hal.inria.fr/hal-02121389

[17] L. Duroyon, F. Goasdoué, I. Manolescu.*A Linked Data Model for Facts, Statements and Beliefs*, in "International Workshop on Misinformation, Computational Fact-Checking and Credible Web", San Francisco, United States, WWW '19 Companion - Proceedings of the 2019 World Wide Web Conference, May 2019 [*DOI :* 10.1145/3308560.3316737], https://hal.inria.fr/hal-02057980

[18] A. Ghazimatin, O. Balalau, R. Saha, G. Weikum.*PRINCE: Provider-side Interpretability with Counterfactual Explanations in Recommender Systems*, in "The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM'20)", Houston, Texas, United States, February 2020, https://hal.inria.fr/hal-02433443

[19] F. Goasdoué, P. Guzewicz, I. Manolescu.*Incremental structural summarization of RDF graphs*, in "EDBT 2019 - 22nd International Conference on Extending Database Technology", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-01978784

[20] P. Guzewicz, I. Manolescu.*Parallel Quotient Summarization of RDF Graphs*, in "SBD 2019 - International Workshop on Semantic Big Data", Amsterdam, Netherlands, June 2019 [*DOI :* 10.1145/3323878.3325809], https://hal.inria.fr/hal-02106521

[21] H. Kondylakis, D. Kotzinos, I. Manolescu.*RDF graph summarization: principles, techniques and applications (tutorial)*, in "EDBT/ICDT 2019 - 22nd International Conference on Extending Database Technology - Joint Conference", Lisbonne, Portugal, March 2019, https://hal.inria.fr/hal-02081474

[22] I. Manolescu, M. Mazuran.*Speeding up RDF aggregate discovery through sampling*, in "BigVis 2019 - 2nd International Workshop on Big Data Visual Exploration and Analytics", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-02065993

[23] L. Tanca, D. Azzalini, F. Azzalini, M. Mazuran.*Tracking the Evolution of Financial Time Series Clusters*, in "DSMM 2019 - 5th Workshop on Data Science for Macro-modeling with Financial and Economic Datasets", Amsterdam, Netherlands, June 2019, https://hal.inria.fr/hal-02191810

National Conferences with Proceeding

[24] D. Azzalini, F. Azzalini, M. Mazuran, L. Tanca.*Evolution of Financial Time Series Clusters (Discussion Paper)*, in "SEBD 2019 - 27th Italian Symposium on Advanced Database Systems", Castiglione della Pescaia (Grosseto), Italy, June 2019, https://hal.inria.fr/hal-02191794

Research Reports

[25] M. BURON, F. GOASDOUÉ, I. MANOLESCU, M.-L. MUGNIER.*Ontology-Based RDF Integration of Heterogeneous Data*, LIX, Ecole polytechnique ; Inria Saclay, August 2019, https://hal.inria.fr/hal-02266517

Other Publications

[26] E. HUANG, L. D. PALMA, L. CETINSOY, Y. DIAO, A. LIU.*AIDEme: An active learning based system for interactive exploration of large datasets*, December 2019, NeurIPS 2019, https://hal.inria.fr/hal-02430750

[27] I. MANOLESCU.*Computational fact-checking: Problems, state of the art and perspectives*, January 2019, 19e Conférence Francophone sur l'Extraction et Gestion de Connaissances (EGC), https://hal.inria.fr/hal-01995318

[28] I. MANOLESCU.*Journalistic Dataspaces: Data Management for Journalism and Fact-Checking (Keynote Talk)*, March 2019, EDBT/ICDT 2019 Joint Conference, https://hal.inria.fr/hal-02081430

Project-Team CELESTE

mathematical statistics and learning

IN COLLABORATION WITH: Laboratoire de mathématiques d'Orsay de l'Université de Paris-Sud (LMO)

IN PARTNERSHIP WITH:
CNRS

Université Paris-Sud (Paris 11)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team CELESTE

Creation of the Project-Team: 2019 June 01

Keywords:

Computer Science and Digital Science:

A3.1.1. - Modeling, representation
A3.1.8. - Big data (production, storage, transfer)
A3.3. - Data and knowledge analysis
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.3. - Reinforcement learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.7. - Kernel methods
A3.5.1. - Analysis of large graphs
A5.9.2. - Estimation, modeling
A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.4. - Statistical methods
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A6.3.3. - Data processing
A6.3.4. - Model reduction
A9.2. - Machine learning

Other Research Topics and Application Domains:

B1.1.4. - Genetics and genomics
B1.1.7. - Bioinformatics
B2.2.4. - Infectious diseases, Virology
B2.3. - Epidemiology
B2.4.1. - Pharmaco kinetics and dynamics
B3.4. - Risks
B4. - Energy
B4.4. - Energy delivery
B4.5. - Energy consumption
B5.2.1. - Road vehicles
B5.2.2. - Railway
B5.2.3. - Aviation
B5.5. - Materials
B5.9. - Industrial maintenance

B7.1. - Traffic management
B7.1.1. - Pedestrian traffic and crowds
B9.5.2. - Mathematics
B9.8. - Reproducibility

1. Team, Visitors, External Collaborators

Research Scientists
Kevin Bleakley [Inria, Researcher]
Gilles Celeux [Inria, Emeritus]
Matthieu Lerasle [CNRS, Researcher]
Gilles Stoltz [CNRS, Researcher, HDR]

Faculty Members
Sylvain Arlot [Team leader, Univ Paris-Sud, Professor]
Alexandre Janon [Univ Paris-Sud, Associate Professor]
Christophe Giraud [Univ Paris-Sud, Professor]
Christine Keribin [Univ Paris-Sud, Associate Professor, HDR]
Pascal Massart [Univ Paris-Sud, Professor]
Marie-Anne Poursat [Univ Paris-Sud, Associate Professor]
Patrick Pamphile [Univ Paris-Sud, Associate Professor]

PhD Students
Yvenn Amara Ouali [Univ Paris-Saclay, from Oct 2019]
Geoffrey Chinot [Univ Paris-Sud]
Olivier Coudray [PSA, from Oct 2019]
Remi Coulaud [SNCF, from Oct 2019]
Solenne Gaucher [Univ Paris-Sud]
Benjamin Goehry [Univ Paris-Sud]
Hedi Hadiji [Ministère de l'Enseignement Supérieur et de la Recherche, from Jun 2019]
Antoine Havet-Morel [Ecole polytechnique, until Aug 2019]
Malo Huard [Univ Paris-Sud, until May 2019]
Yann Issartel [Univ Paris-Saclay]
Perrine Lacroix [Univ Paris-Sud, from Oct 2019]
Guillaume Maillard [Univ Paris-Sud]
Timothee Mathieu [Ecole Normale Supérieure Cachan]
Minh-Lien Nguyen [Univ Paris-Sud]
El Mehdi Saad [Univ Paris-Sud, from Sep 2019]
Solene Thepaut [Univ Paris-Sud]

Post-Doctoral Fellow
Evgenii Chzhen [Univ Paris-Sud, from Oct 2019]

2. Overall Objectives

2.1. Mathematical statistics and learning

Data science – a vast field that includes statistics, machine learning, signal processing, data visualization, and databases – has become front-page news due to its ever-increasing impact on society, over and above the important role it already played in science over the last few decades. Within data science, the statistical community has long-term experience in how to infer knowledge from data, based on solid mathematical foundations. The more recent field of machine learning has also made important progress by combining statistics and optimization, with a fresh point of view that originates in applications where prediction is more important than building models.

The CELESTE project-team is positioned at the interface between statistics and machine learning. We are statisticians in a mathematics department, with strong mathematical backgrounds behind us, interested in interactions between theory, algorithms and applications. Indeed, applications are the source of many of our interesting theoretical problems, while the theory we develop plays a key role in (i) understanding how and why successful statistical learning algorithms work – hence improving them – and (ii) building new algorithms upon mathematical statistics-based foundations

In the theoretical and methodological domains, CELESTE aims to analyze statistical learning algorithms – especially those which are most used in practice – with our mathematical statistics point of view, and develop new learning algorithms based upon our mathematical statistics skills.

A key ingredient in our research program is connecting our theoretical and methodological results with (a great number of) real-world applications. Indeed, CELESTE members work in many domains, including – but not limited to – neglected tropical diseases, pharmacovigilance, high-dimensional transcriptomic analysis, and energy and the environment.

3. Research Program

3.1. General presentation

Our objectives correspond to four major challenges of machine learning where mathematical statistics have a key role. First, any machine learning procedure depends on hyperparameters that must be chosen, and many procedures are available for any given learning problem: both are an estimator selection problem. Second, with high-dimensional and/or large data, the computational complexity of algorithms must be taken into account differently, leading to possible trade-offs between statistical accuracy and complexity, for machine learning procedures themselves as well as for estimator selection procedures. Third, real data are almost always corrupted partially, making it necessary to provide learning (and estimator selection) procedures that are robust to outliers and heavy tails, while being able to handle large datasets. Fourth, science currently faces a reproducibility crisis, making it necessary to provide statistical inference tools (p-values, confidence regions) for assessing the significance of the output of any learning algorithm (including the tuning of its hyperparameters), in a computationally efficient way.

3.2. Estimator selection

An important goal of CELESTE is to build and study procedures that can deal with general estimators (especially those actually used in practice, which often rely on some optimization algorithm), such as cross-validation and Lepski's method. In order to be practical, estimator selection procedures must be fully data-driven (that is, not relying on any unknown quantity), computationally tractable (especially in the high-dimensional setting, for which specific procedures must be developed) and robust to outliers (since most real data sets include a few outliers). CELESTE aims at providing a precise theoretical analysis (for new and existing popular estimator selection procedures), that explains as well as possible their observed behaviour in practice.

3.3. Relating statistical accuracy to computational complexity

When several learning algorithms are available, with increasing computational complexity and statistical performance, which one should be used, given the amount of data and the computational power available? This problem has emerged as a key question induced by the challenge of analyzing large amounts of data – the "big data" challenge. CELESTE wants to tackle the major challenge of understanding the time-accuracy trade-off, which requires providing new statistical analyses of machine learning procedures – as they are done in practice, including optimization algorithms – that are *precise enough* in order to account for differences of performance observed in practice, leading to general conclusions that can be trusted more generally. For instance, we study the performance of ensemble methods combined with subsampling, which is a common strategy for handling big data; examples include random forests and median-of-means algorithms.

3.4. Robustness to outliers and heavy tails (with tractable algorithms)

The classical theory of robustness in statistics has recently received a lot of attention in the machine learning community. The reason is simple: large datasets are easily corrupted, due to – for instance – storage and transmission issues, and most learning algorithms are highly sensitive to dataset corruption. For example, the lasso can be completely misled by the presence of even a single outlier in a dataset. A major challenge in robust learning is to provide computationally tractable estimators with optimal subgaussian guarantees. A second important challenge in robust learning is to deal with datasets where every (x_i, y_i) is slightly corrupted. In large-dimensional data, every single data point x_i is likely to have several corrupted coordinates, and no estimator currently has strong theoretical guarantees for such data. A third important challenge is that of robust estimator selection or aggregation. Even if several robust estimators can be built, the final aggregation or selection step in a user's routine is usually based on empirical means. This is not robust, and may damage the global performance of the procedure. Instead, we can consider more sophisticated types of aggregation of the base robust estimators built so far. A convenient framework to do so is called adversarial learning (also known as: prediction of individual sequences). Here, data is not assumed to be stochastic, and it could even be chosen by an adversary.

3.5. Statistical inference: (multiple) tests and confidence regions (including post-selection)

CELESTE considers the problems of quantifying the uncertainty of predictions or estimations (thanks to confidence intervals) and of providing significance levels (p-values, corrected for multiplicity if needed) for each "discovery" made by a learning algorithm. This is an important practical issue when performing feature selection – one then speaks of post-selection inference – change-point detection or outlier detection, to name but a few. We tackle it in particular through a collaboration with the Parietal team (Inria Saclay) and LBBE (CNRS), with applications in neuroimaging and genomics.

4. Application Domains

4.1. Neglected tropical diseases

CELESTE collaborates with Anavaj Sakuntabhai and Philippe Dussart (Pasteur Institute) on predicting dengue severity using only low-dimensional clinical data obtained at hospital arrival. Further collaborations are underway in dengue fever and encephalitis with researchers at the Pasteur Institute, including with Jean-David Pommier.

4.2. Pharmacovigilance

In pharmacovigilance, the goal is to detect, as soon as possible, potential associations between certain drugs and adverse effects, which appeared after the authorized marketing of these drugs. Preceding works showed the importance of defining an adapted methodology to deal with the size of the individual data (around 250000 reports, 2000 drugs, 4000 adverse effects) and their sparsity. We will explore several aspects from software point of view to statistical strategies as sub-sampling.

4.3. Electricity load consumption: forecasting and control

CELESTE has a long-term collaboration with EDF R&D about electricity consumption. An important problem is to forecast consumption. We currently work on an approach involving back and forth disaggregation (of the total consumption into the consumptions of well-chosen groups/regions) and aggregation of local estimates. We also work on consumption control by price incentives sent to specific users (volunteers), seeing it as a bandit problem.

4.4. Reliability

Collected product lifetime data is often non-homogeneous, affected by production variability and differing real-world usage. Usually, this variability is not controlled or observed in any way, but needs to be taken into account for reliability analysis. Latent structure models are flexible models commonly used to model unobservable causes of variability.

CELESTE currently collaborates with PSA Group. To dimension its vehicles, the PSA Group uses a reliability design method called Strength-Stress, which takes into consideration both the statistical distribution of part strength and the statistical distribution of customer load (called Stress). In order to minimize the risk of in-service failure, the probability that a severe customer will encounter a weak part must be quantified. Severity quantification is not simple since vehicle use and driver behaviour can be severe for some types of materials and not for others. The aim of the study is then to define a new and richer notion of the severity from the PSA databases resulting either from tests or client usages. This will lead to a more robust and accurate parts dimensioning method.

4.5. Spectroscopic imaging analysis of ancient materials

Ancient materials, encountered in archaeology and paleontology are often complex, heterogeneous and poorly characterized before physico-chemical analysis. A popular technique is to gather as much physico-chemical information as possible, is spectro-microscopy or spectral imaging, where a full spectra, made of more than a thousand samples, is measured for each pixel. The produced data is tensorial with two or three spatial dimensions and one or more spectral dimensions, and requires the combination of an "image" approach with a "curve analysis" approach. Since 2010 CELESTE (previously SELECT) collaborates with Serge Cohen (IPANEMA) on clustering problems, taking spatial constraints into account.

4.6. Forecast of dwell time during train parking at stations

This is a Cifre PhD in collaboration with SNCF.

One of the factors in the punctuality of trains in dense areas (and management crisis in the event of an incident on a line) is the respect of both the travel time between two stations and the parking time in a station. These depend, among other things, on the train, its mission, the schedule, the instantaneous charge, and the configuration of the platform or station. Preliminary internal studies at SNCF have shown that the problem is complex. From a dataset concerning line E of the Transilien in Paris, we will address prediction (machine learning) and modeling (statistics): (1) construct a model of station-hours, station-hours-type of train, by example using co-clustering techniques; (2) study the correlations between the number of passengers (load), up and down flows, and parking times, and possibly other variables to be defined; (3) model the flows or loads (within the same station, or the same train) as a stochastic process; (4) develop a realistic digital simulator of passenger flows and test different scenarios of incidents and resolution, in order to propose effective solutions.

4.7. Fatigue aided-design

This is a Cifre PhD in collaboration with PSA.

The digitalization of design is at the heart of the processes of automotive manufacturers departments, to enable them to reduce costs and development time. This also applies to reliability studies of certain components of the chassis of a vehicle, and the will is to drastically reduce the number of physical tests to tend towards an almost entirely digital design having only one phase of validation. Deterministic models, although developed from detailed design drawings, can predict behaviors different from those observed on the structure during testing. These deviations can be due to the more or less faithful discretization of the geometry, the uncertainties on some parameters of the model (such as the properties of the materials, the boundary conditions), or the random loadings undergone by the structure (Beck and Katafygiotis, 1998). It is important to make available new methods in addition to the classical finite element (FE) deterministic modeling, to enable the exploitation of the accumulated data over the years for all the projects: computation results, measurements and test data.

One of the objectives of this project is to propose a probabilistic modeling of the behavior of a structure starting from a FE model, taking into account the non assignable fluctuations of the model, in order to define a probabilistic criterion of rupture and its margins of confidence. The following three steps are envisaged: (1) Define relevant prior information using business experience feedback (REX) and use a Bayesian estimation to calibrate the parameters. This REX is consequent and will require advanced statistical processing of machine learning, and in particular in clustering to identify similarities or similar patterns among several models. The estimation will use Bayesian non-iterative methods (Celeux and Pamphile, 2019), which are less expensive and less unstable than conventional methods. This will test their effectiveness in this context. (2) Select important parameters (physical or modeling). (3) Define a probabilistic criterion of coaxial fatigue taking into account both the random behavior of the structure and the material (Fouchereau et al., 2014) extending the existing deterministic criteria (Dang-Van, 1993).

5. New Software and Platforms

5.1. BlockCluster

Block Clustering

KEYWORDS: Statistic analysis - Clustering package

SCIENTIFIC DESCRIPTION: Simultaneous clustering of rows and columns, usually designated by biclustering, co-clustering or block clustering, is an important technique in two way data analysis. It consists of estimating a mixture model which takes into account the block clustering problem on both the individual and variables sets. The blockcluster package provides a bridge between the C++ core library and the R statistical computing environment. This package allows to co-cluster binary, contingency, continuous and categorical data-sets. It also provides utility functions to visualize the results. This package may be useful for various applications in fields of Data mining, Information retrieval, Biology, computer vision and many more.

FUNCTIONAL DESCRIPTION: BlockCluster is an R package for co-clustering of binary, contingency and continuous data based on mixture models.

RELEASE FUNCTIONAL DESCRIPTION: Initialization strategy enhanced

- Participants: Christophe Biernacki, Gilles Celeux, Parmeet Bhatia, Serge Iovleff, Vincent Brault and Vincent Kubicki
- Partner: Université de Technologie de Compiègne
- Contact: Serge Iovleff
- URL: http://cran.r-project.org/web/packages/blockcluster/index.html

5.2. MASSICCC

Massive Clustering with Cloud Computing

KEYWORDS: Statistic analysis - Big data - Machine learning - Web Application

SCIENTIFIC DESCRIPTION: The web application let users use several software packages developed by Inria directly in a web browser. Mixmod is a classification library for continuous and categorical data. MixtComp allows for missing data and a larger choice of data types. BlockCluster is a library for co-clustering of data. When using the web application, the user can first upload a data set, then configure a job using one of the libraries mentioned and start the execution of the job on a cluster. The results are then displayed directly in the browser allowing for rapid understanding and interactive visualisation.

FUNCTIONAL DESCRIPTION: The MASSICCC web application offers a simple and dynamic interface for analysing heterogeneous data with a web browser. Various software packages for statistical analysis are available (Mixmod, MixtComp, BlockCluster) which allow for supervised and supervised classification of large data sets.

- Contact: Christophe Biernacki
- URL: https://massiccc.lille.inria.fr

5.3. Mixmod

Many-purpose software for data mining and statistical learning

KEYWORDS: Data mining - Classification - Mixed data - Data modeling - Big data

FUNCTIONAL DESCRIPTION: Mixmod is a free toolbox for data mining and statistical learning designed for large and highdimensional data sets. Mixmod provides reliable estimation algorithms and relevant model selection criteria.

It has been successfully applied to marketing, credit scoring, epidemiology, genomics and reliability among other domains. Its particularity is to propose a model-based approach leading to a lot of methods for classification and clustering.

Mixmod allows to assess the stability of the results with simple and thorough scores. It provides an easy-to-use graphical user interface (mixmodGUI) and functions for the R (Rmixmod) and Matlab (mixmodForMatlab) environments.

- Participants: Benjamin Auder, Christophe Biernacki, Florent Langrognet, Gérard Govaert, Gilles Celeux, Remi Lebret and Serge Iovleff
- Partners: CNRS - Université Lille 1 - LIFL - Laboratoire Paul Painlevé - HEUDIASYC - LMB
- Contact: Gilles Celeux
- URL: http://www.mixmod.org

6. New Results

6.1. Minimal penalty algorithms for model selection

Birgé and Massart proposed in 2001 the slope heuristics as a way to choose optimally from data an unknown multiplicative constant in front of a penalty. It is built upon the notion of minimal penalty, and it has been generalized since to some “minimal-penalty algorithms”. The survey [3] by S. Arlot reviews the theoretical results obtained for such algorithms, with a self-contained proof in the simplest framework, precise proof ideas for further generalizations, and a few new results. Explicit connections are made with residual-variance estimators —with an original contribution on this topic, showing that for this task the slope heuristics performs almost as well as a residual-based estimator with the best model choice— and some classical algorithms such as L-curve or elbow heuristics, Mallows' C_p, and Akaike's FPE. Practical issues are also addressed, including two new practical definitions of minimal-penalty algorithms that are compared on synthetic data to previously-proposed definitions. Finally, several conjectures and open problems are suggested as future research directions. This extensive survey is followed by a discussion by 13 authors, and a rejoinder in which another original result is proved (theoretical validation of the slope heuristics when all models in the collection are biased).

6.2. Kernel change-point detection

In collaboration with A. Celisse and Z. Harchaoui, S. Arlot worked on the change-point problem with data belonging to a general set. They built a penalty for choosing the number of change-points in the kernel-based method of Harchaoui and Cappé (2007). This penalty generalizes the one proposed by Lebarbier (2005) for one-dimensional signals. They prove in [4] a non-asymptotic oracle inequality for the proposed method, thanks to a new concentration result for some function of Hilbert-space valued random variables. Experiments on synthetic and real data illustrate the accuracy of our method, showing that it can detect changes in the whole distribution of data, even when the mean and variance are constant. This method has since been used successfully by several authors in various applied contexts.

6.3. A probabilistic method to characterize genomic alterations of tumors

Characterizing the genomic copy number alterations (CNA) in cancer is of major importance in order to develop personalized medicine. Single nucleotide polymorphism (SNP) arrays are still in use to measure CNA profiles. Among the methods for SNP-array analysis, the Genome Alteration Print (GAP) by Popova et al, based on a preliminary segmentation of SNP-array profiles, uses a deterministic approach to infer the absolute copy numbers profile. C. Keribin with Y. Liu, Y. Rozenholch and T. Popova developed a probabilistic model in [9] for GAP and define a Gaussian mixture model where centers are constrained to belong to a frame depending on unknown parameters such as the proportion of normal tissue. The estimation is performed using an expectation-maximization (EM) algorithm to recover the parameters characterizing the genomic alterations as well as the most probable copy number change of each segment and the unknown proportion of normal tissue. The tumor ploidy can be deduced from penalized model selection criterion. The model is tested on simulated and real data.

Surprisingly, the BIC selection criterion cannot recover the actual ploidy in the real data sets as slope heuristics do, even though all models are wrong. C. Keribin, in a discussion of S. Arlot's survey, gave some arguments to explain these behaviors [8].

6.4. New results for stochastic bandits

M. Brégère and G. Stoltz, in collaboration with P. Gaillard (Sierra team) and Y. Goude (EDF), provided a methodology in [5] based on a modeling by linear bandits, for managing (influencing) electricity consumption by sending tariff incentives. The main result is the very modeling of the problem: consumption is modeled as a generalized additive model based on the probabilistic allocation of tariffs picked and on the context (given by the type of day, hour of the day, weather conditions, etc.). Mathematical results are, on the other hand, direct extensions of earlier results for the LinUCB algorithm (see Li et al., 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011). Simulations on realistic data are provided: for bandit algorithms, one needs a data simulator, which we created based on an open data set consisting of households in London.

A second important result was obtained by H. Hadiji: he characterized the cost of adaptation to the unknown (Höderian) smoothness payoff functions in continuum-armed bandits [14]. He first rewrote and slightly extended the regret lower bounds exhibited by Locatelli and Carpentier (2018), and then exhibited an algorithm with matching regret upper bounds. This algorithm, unlike virtually all previous algorithms in X-armed bandits, which zoomed in as time passes, zooms out as time passes. This solves a problem that had been open for several years.

Also, H. Hadiji and G. Stoltz, in collaboration with P. Ménard (SequeL team) and A. Garivier, submitted a revised version of their results of simultaneous optimality (from both a distribution-dependent and a distribution-free viewpoints) for a variant of the KL-UCB algorithm in the case of vanilla K-armed stochastic bandits [22].

6.5. Robust risk minimization for machine learning

In collaboration with S. Minsker (USC), T. Mathieu worked on obtaining new excess risk bounds in robust empirical risk minimization. The method proposed in [29] is inspired from the robust risk minimization procedure using median-of-means estimators in Lecué, Lerasle and Mathieu (2018). The obtained excess risk are faster than the so-called "slow rate of convergence" obtained for the minimization procedure in Lecué, Lerasle and Mathieu (2018) and a slightly modified procedure achieves a minimax rate of convergence under low moment assumptions. Experiments on synthetic corrupted data and real dataset illustrate the accuracy of the method showing high performance in classification and regression tasks in a corrupted setting.

6.6. Optimal pair-matching

The sequential pair-matching problem appears in many applications (in particular for the internet) where one wants to discover, sequentially, good matches between pairs of individuals, for a given budget. C. Giraud,

Y. Issartel, L. Lehéricy and M. Lerasle propose a formulation of this problem as a special bandit problem on graphs [23]. Formally, the set of individuals is represented by the nodes of a graph where the edges, unobserved at first, represent the potential good matches. The algorithm queries pairs of nodes and observes the presence/absence of edges. Its goal is to discover as many edges as possible with a fixed budget of queries. Pair-matching is a particular instance of multi-armed bandit problem in which the arms are pairs of individuals and the rewards are edges linking these pairs. This bandit problem is non-standard though, as each arm can only be played once.

Given this last constraint, sublinear regret can be expected only if the graph has some underlying structure. C. Giraud, Y. Issartel, L. Lehéricy and M. Lerasle show in [23] that sublinear regret is achievable in the case where the graph is generated according to a Stochastic Block Model (SBM) with two communities. Optimal regret bounds are computed for this pair-matching problem. They exhibit a phase transition related to the Kesten-Stigund threshold for community detection in SBMs. In practice, it is meaningful to constrain each node to be sampled less than a given amount of times, in order to avoid concentration of queries on a set of individuals. This setting is more challenging both on the statistical side and the algorithmic side. Optimal rates are also derived in this context, exhibiting how the regret deteriorates with this constraint.

6.7. Minimax estimation of network complexity in graphon model

In network analysis, the graphon model has attracted a lot of attention as a non-parametric model with some universal properties. However, this approach suffers from interpretability and identifiability issues in practice. A first solution to this problem was obtained by Y. Issartel: he introduces an identifiable and interpretable functional of the graphon, which measures the complexity of network [25]. It has simple interpretations on popular examples of random graphs: it matches the number of communities in stochastic block models; the dimension of the Euclidean space in random geometric graphs; the regularity of the link function in Hölder graphons. He also provides an estimation procedure of this complexity that is minimax optimal.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR

Sylvain Arlot and Matthieu Lerasle are part of the ANR grant FAST-BIG (Efficient Statistical Testing for high-dimensional Models: application to Brain Imaging and Genetics), which is lead by Bertrand Thirion (Inria Saclay, Parietal).

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

8.1.1.1. Member of the Organizing Committees

Semaine SEME - Orsay 14/01-18/01-2019, https://www.math.u-psud.fr/seme2019/index.php. G. Stoltz was part of the organizing and scientific committees. C. Keribin was part of the scientific committee.

8.1.2. Scientific Events: Selection

8.1.2.1. Member of the Conference Program Committees

- S. Arlot was member of the steering committee of the 4th Junior Conference on Data Science and Engineering at Paris-Saclay (Sept. 2019), CentraleSupélec, Paris-Saclay campus, Gif-sur-Yvette.
- C. Keribin is member of steering committee of the the bi-montly Seminar of probability and statistics (Laboratoire de mathématiques d'Orsay)
- C. Giraud is co-organiser (with E. Kuhn) of the conference StatMathAppli at Fréjus (1-6 september 2019)
- C. Giraud is local member of scientific committee of the Institut Pascal (year around program)

8.1.2.2. Reviewer

We performed many reviews for various international conferences.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

S. Arlot is associate editor of Annales de l'Institut Henri Poincaré B – Probability and Statistics.

8.1.3.2. Reviewer - Reviewing Activities

We performed many reviews for various international journals.

8.1.4. Invited Talks

- S. Arlot, Changepoint Workshop, Nov. 2019, Institut des Systèmes Complexes, Paris.
- S. Arlot, Thematic week "Data and Analytics for Short-Term Operations", Feb. 2019, Isaac Newton Institute for Mathematical Science, Cambridge, UK.
- C. Keribin, Some asymptotic properties of model selection criteria in the latent block model. 12th Scientific meeting CLADAG 2019, Cassino (Italie), September 11 – 13, 2019
- C. Giraud, Community detection and sequential learning, IOPS 2019, Bordeaux, June 19-21, 2019
- C. Giraud, Sequential learning in random graph, Genova, May 2019.
- P. Pamphile, Maintenance cost forecasting for a feet of vehicles, IMDR, Paris, April 2019

8.1.5. Research Administration

S. Arlot coordinates the math-AI (mathematics for artificial intelligence) program of the Labex Mathématique Hadamard and is member of the executive comittee of Fondation Mathématique Jacques Hadamard (FMJH).
S. Arlot is member of the steering committee of the Paris-Saclay Center for Data Science.
S. Arlot is member of the prefiguration group of the Computer Science Graduate School of University Paris-Saclay.
P. Massart is Director of the Fondation Mathématique Jacques Hadamard (FMJH).
C. Giraud has coordinated the math-SV (mathematics for life science) program of the Labex Mathématique Hadamard and is member of the executive comittee of Fondation Mathématique Jacques Hadamard (FMJH).
C. Giraud is member of the scientific committee of the Labex IRMIA (Strasbourg)
C. Giraud is local member of the scientific committee of the Pascal Institute (Saclay)
C. Giraud is member of the prefiguration group of the Mathematics Graduate School of University Paris-Saclay.
C. Giraud is member of the steering committee of the Mathematics Modelisation and Biodiversity chair.
C. Giraud is in charge of the whole master program in Mathematics of Paris Saclay.
C. Giraud is in charge of the Statistics and Machine Learning track in the master program in Mathematics of Paris Saclay.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence: S. Arlot, Probability and Statistics, 68h, L2, Université Paris-Sud
Master: S. Arlot, Statistical learning and resampling, 30h, M2, Université Paris-Sud
Master: S. Arlot, Probability and Statistics M2 seminar, 30h, M2, Université Paris-Sud
Master: S. Arlot, Preparation to French mathematics agrégation (statistics), 50h, M2, Université Paris-Sud
Master: C. Giraud, High-Dimensional Statistics, 45h, M2, Université Paris-Sud
Master: C. Giraud, Theoretical Guidelines in Data Analysis, 45h, M2, Université Paris-Sud
Master: C. Giraud, Lecture Group, 25h, M2, Université Paris-Sud
Master: C. Giraud, Mathematics for AI, 75h, M1, Université Paris-Sud

8.2.2. Supervision

PhD in progress: Guillaume Maillard, Aggregated cross-validation, started Sept. 2016, co-advised by S. Arlot and M. Lerasle
PhD in progress: El Mehdi Saad, Interactions between statistical and computational aspects in machine learning, started Sept. 2019, co-advised by S. Arlot and G. Blanchard
PhD in progress: Tuan-Binh Nguyen, Efficient Statistical Testing for High-Dimensional Models, co-advised by S. Arlot and B. Thirion
PhD in progress: Rémi Coulaud, Forecast of dwell time during train parking at station, started Oct. 2019, co-advised by G. Stoltz and C. Keribin, Cifre with SNCF
PhD in progress: Olivier Coudray, Fatigue aided-design, started Nov. 2019, co-advised by C. Keribin and P. Pamphile, Cifre with PSA
PhD: Solène Thépaut, Problèmes de clustering liés à la synchronie en écologie, Université Paris Saclay, Dec. 2019, C. Giraud.
PhD: Théophile Olivier, Le role de la diversité et des perturbations environnementales sur la stabilité temporelle des communautés animales en milieu naturel, Museum National Histoire Naturelle, Sep. 2019, co-advised by E. Porcher and C. Giraud.
PhD in progress: Yann Issartel, Non-parametric estimation in random networks, started Sep. 2017, C. Giraud.
PhD in progress: Solenne Gaucher, Sequential learning in random networks, started Sep. 2018, C. Giraud.

8.2.3. Juries

S. Arlot: referee for the HdR of Servane Gey, Université Paris Descartes, 07/02/2019.
S. Arlot: member of the PhD committee of Solène Thepaut, Université Paris-Sud, 06/12/2019.
C. Giraud: many HDR and PhD juries as referee or member of the committee

8.3. Popularization

8.3.1. Interventions

- Public exhibitions: S. Arlot is member of the steering committee of a general-audience exhibition about artificial intelligence, that is co-organized by Fermat Science (Toulouse), Institut Henri Poincaré (IHP, Paris) and Maison des Mathématiques et de l'Informatique (MMI, Lyon).

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] C. KERIBIN.*From clustering to co-clustering : a model based approach*, Université Paris Sud XI, November 2019, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-02397429

Articles in International Peer-Reviewed Journal

[2] S. ARLOT. *Minimal penalties and the slope heuristics: a survey*, in "Journal de la Socie´te´ Française de Statistique", October 2019, vol. 160, n^{o} 3, p. 1-106, https://arxiv.org/abs/1901.07277 , https://hal.archives-ouvertes.fr/hal-01989167

[3] S. ARLOT. *Rejoinder on: Minimal penalties and the slope heuristics: a survey*, in "Journal de la Socie´te´ Française de Statistique", October 2019, vol. 160, n^{o} 3, p. 158-168, https://arxiv.org/abs/1909.13499 , https://hal.archives-ouvertes.fr/hal-02300688

[4] S. ARLOT, A. CELISSE, Z. HARCHAOUI. *A Kernel Multiple Change-point Algorithm via Model Selection*, in "Journal of Machine Learning Research", December 2019, vol. 20, n^{o} 162, p. 1–56, https://arxiv.org/abs/1202.3878 , https://hal.archives-ouvertes.fr/hal-00671174

[5] M. BRÉGÈRE, P. GAILLARD, Y. GOUDE, G. STOLTZ. *Target Tracking for Contextual Bandits: Application to Demand Side Management*, in "Proceedings of Machine Learning Research", June 2019, vol. 97, p. 754-763, https://arxiv.org/abs/1901.09532 , https://hal.archives-ouvertes.fr/hal-01994144

[6] G. CHINOT, G. LECUÉ, M. LERASLE. *Robust Statistical learning with Lipschitz and convex loss functions*, in "Probability Theory and Related Fields", July 2019, https://arxiv.org/abs/1810.01090 [*DOI :* 10.1007_s00440-019-00931-3], https://hal.archives-ouvertes.fr/hal-01923033

[7] A. HAVET, M. LERASLE, É. MOULINES. *Density estimation for RWRE*, in "Mathematical Methods of Statistics", March 2019, https://arxiv.org/abs/1806.05839 [*DOI :* 10.3103/S1066530719010022], https://hal.archives-ouvertes.fr/hal-01815990

[8] C. KERIBIN. *A note on BIC and the slope heuristic*, in "Journal de la Socie´te´ Française de Statistique", 2019, https://hal.archives-ouvertes.fr/hal-02391310

[9] C. KERIBIN, Y. LIU, T. POPOVA, Y. ROZENHOLC. *A mixture model to characterize genomic alterations of tumors*, in "Journal de la Socie´te´ Française de Statistique", 2019, https://hal.archives-ouvertes.fr/hal-02391289

[10] G. LECUÉ, M. LERASLE. *Robust machine learning by median-of-means : theory and practice*, in "Annals of Statistics", February 2019, https://arxiv.org/abs/1711.10306 - 48 pages, 6 figures, https://hal.archives-ouvertes.fr/hal-01923036

Invited Conferences

[11] C. BIERNACKI, G. CELEUX, J. JOSSE, F. LAPORTE. *Dealing with missing data in model-based clustering through a MNAR model*, in "CRoNos & MDA 2019 - Meeting and Workshop on Multivariate Data Analysis and Software", Limassol, Cyprus, April 2019, https://hal.inria.fr/hal-02103347

International Conferences with Proceedings

[12] M. LERASLE, Z. SZABÓ, T. MATHIEU, G. LECUÉ. *MONK – Outlier-Robust Mean Embedding Estimation by Median-of-Means*, in "ICML 2019 - 36th International Conference on Machine Learning", Long Beach, United States, Proceedings of Machine Learning Research, June 2019, https://hal.archives-ouvertes.fr/hal-01705881

Conferences without Proceedings

[13] E. CHZHEN, C. DENIS, M. HEBIRI, L. ONETO, M. PONTIL.*Leveraging Labeled and Unlabeled Data for Consistent Fair Binary Classification*, in "NeurIPS", Vancouver, Canada, December 2019, https://arxiv.org/abs/1906.05082 , https://hal-upec-upem.archives-ouvertes.fr/hal-02150662

[14] H. HADIJI.*Polynomial Cost of Adaptation for X -Armed Bandits*, in "Thirty-third Conference on Neural Information Processing Systems", Vancouver, France, December 2019, https://arxiv.org/abs/1905.10221 , https://hal.archives-ouvertes.fr/hal-02138492

[15] C. KERIBIN, C. BIERNACKI.*Co-clustering: model based or model free approaches*, in "62nd ISI World Statistics Congress 2019", Kuala Lumpur, Malaysia, August 2019, https://hal.archives-ouvertes.fr/hal-02399031

[16] C. KERIBIN, C. BIERNACKI.*Le modèle des blocs latents, une méthode régularisée pour la classification en grande dimension*, in "JdS 2019 - 51èmes Journées de Statistique de la SFdS", Nancy, France, June 2019, https://hal.archives-ouvertes.fr/hal-02391379

[17] C. KERIBIN.*Some Asymptotic Properties of Model Selection Criteria in the Matent Block Model*, in "CLADAG 2019 - 12th Scientific Meeting Classification and Data Analysis Group", Cassino, Italy, September 2019, https://hal.archives-ouvertes.fr/hal-02391398

[18] F. LAPORTE, C. BIERNACKI, G. CELEUX, J. JOSSE.*Modèles de classification non supervisée avec données manquantes non au hasard*, in "51e journées de statistique", Nancy, France, June 2019, https://hal.archives-ouvertes.fr/hal-02398984

Other Publications

[19] G. CELEUX, P. PAMPHILE.*Estimating parameters of the Weibull Competing Risk model with Masked Causes and Heavily Censored Data*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02410489

[20] G. CHINOT, G. LECUÉ, M. LERASLE.*Robust high dimensional learning for Lipschitz and convex losses*, June 2019, https://arxiv.org/abs/1905.04281 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02159943

[21] F. DUCROS, P. PAMPHILE.*Maintenance cost forecasting for a fleet of vehicles*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02055175

[22] A. GARIVIER, H. HADIJI, P. MENARD, G. STOLTZ.*KL-UCB-switch: optimal regret bounds for stochastic bandits from both a distribution-dependent and a distribution-free viewpoints*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01785705

[23] C. GIRAUD, Y. ISSARTEL, L. LEHÉRICY, M. LERASLE.*Pair Matching: When bandits meet stochastic block model*, June 2019, https://arxiv.org/abs/1905.07342 - 57 pages, https://hal.archives-ouvertes.fr/hal-02159938

[24] A. HAVET, M. LERASLE, É. MOULINES, E. VERNET.*A quantitative Mc Diarmid's inequality for geometrically ergodic Markov chains*, July 2019, https://arxiv.org/abs/1907.02809 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02177452

[25] Y. ISSARTEL.*On the Estimation of Network Complexity: Dimension of Graphons*, December 2019, https://arxiv.org/abs/1909.02900 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02413537

[26] A. JANON.*Global optimization using Sobol indices*, June 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02154121

[27] F. LAPORTE, C. BIERNACKI, G. CELEUX, J. JOSSE.*Model-based clustering with missing not at random data. Missing mechanism*, July 2019, Working Group on Model-Based Clustering Summer Session, Poster, https://hal.archives-ouvertes.fr/hal-02398987

[28] G. MAILLARD, S. ARLOT, M. LERASLE.*Aggregated Hold-Out*, September 2019, https://arxiv.org/abs/1909.04890 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02273193

[29] S. MINSKER, T. MATHIEU.*Excess risk bounds in robust empirical risk minimization*, December 2019, https://arxiv.org/abs/1910.07485 - working paper or preprint [*DOI :* 10.07485], https://hal.archives-ouvertes.fr/hal-02390397

[30] V. ROBERT, Y. VASSEUR, V. BRAULT.*Comparing high dimensional partitions with the Coclustering Adjusted Rand Index*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01524832

Project-Team COMETE

Concurrency, Mobility and Transactions

IN COLLABORATION WITH: Laboratoire d'informatique de l'école polytechnique (LIX)

IN PARTNERSHIP WITH:
CNRS

Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Security and Confidentiality

Table of contents

Project-Team COMETE

Creation of the Project-Team: 2008 January 01

Keywords:

Computer Science and Digital Science:

A2.1.1. - Semantics of programming languages
A2.1.5. - Constraint programming
A2.1.6. - Concurrent programming
A2.1.9. - Synchronous languages
A2.4.1. - Analysis
A2.4.2. - Model-checking
A3.4. - Machine learning and statistics
A4.1. - Threat analysis
A4.5. - Formal methods for security
A4.8. - Privacy-enhancing technologies
A8.6. - Information theory
A9.1. - Knowledge
A9.2. - Machine learning
A9.7. - AI algorithmics
A9.9. - Distributed AI, Multi-agent

Other Research Topics and Application Domains:

B6.1. - Software industry
B6.6. - Embedded systems
B9.5.1. - Computer science
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Catuscia Palamidessi [Team leader, Inria, Senior Researcher]
Frank Valencia [CNRS, Researcher]

External Collaborators

Sayan Biswas [University of Bath, from Oct 2019]
Konstantinos Chatzikokolakis [CNRS]
Noemie Fong [ENS Paris, from Apr 2019 until Aug 2019]
Federica Granese [Univ. of Rome “La Sapienza”, from Jul 2019 until Oct 2019]
Juan Pablo Piantanida [Centrale-Supélec, from Oct 2019]

Technical Staff

Ehab Elsalamouny [Inria, Engineer]

PhD Students

Ganesh Del Grosso Guzman [Inria, PhD Student, from Oct 2019]
Natasha Fernandes [Macquarie Univ., PhD Student]
Federica Granese [Inria, PhD Student, from Nov 2019]
Anna Pazii [École polytechnique, PhD Student]

Santiago Quintero [École polytechnique, PhD Student]
Marco Romanelli [Inria, PhD Student]

Post-Doctoral Fellow
Valentina Castiglioni [Inria, Post-Doctoral Fellow, until Apr 2019]

Visiting Scientists
Yusuke Kawamoto [AIST, March 2019 and Nov–Dec 2019]
Sophia Knight [University of Minesota, May 2019]
Takao Murakami [AIST, March 2019]
Carlos Olarte [Universidade Federal do Rio Grande do Norte, Brazil, Nov 2019]
Carlos Pinzon [Universidad Javeriana of Cali, Nov 2019]
Sergio Ramirez [Universidad Javeriana of Cali, from Nov 2019 until Dec 2019]
Camilo Rocha [Universidad Javeriana of Cali, Jun 2019]
Camilo Rueda [Universidad Javeriana of Cali, from May 2019 until Jun 2019]

Administrative Assistant
Maria Agustina Ronco [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

Our times are characterized by the massive presence of highly *distributed systems* consisting of diverse and specialized devices, forming heterogeneous networks, and providing different services and applications. Revolutionary phenomena such as *social networks* and *cloud computing* are examples of such systems.

In Comète we study emerging concepts of this new era of computing. *Security* and *privacy* are some of the fundamental concerns that arise in this setting. In particular, in the modern digital world the problem of keeping information secret or confidential is exacerbated by orders of magnitude: the frequent interaction between users and electronic devices, and the continuous connection between these devices and the internet, offer malicious agents the opportunity to gather and store huge amount of information, often without the individual even being aware of it. Mobility is an additional source of vulnerability, since tracing may reveal significant information. To avoid these kinds of hazards, *security protocols* and various techniques for privacy protection have been designed. However, the properties that they are supposed to ensure are rather subtle, and, furthermore, it is difficult to foresee all possible expedients that a potential attacker may use. As a consequence, even protocols that seem at first "obviously correct" are later (often years later) found to be prone to attacks.

In addition to the security problems, the problems of correctness, robustness and reliability are made more challenging by the complexity of these systems, since they are highly concurrent and distributed. Despite being based on impressive engineering technologies, they are still prone to faulty behavior due to errors in the software design.

To overcome these drawbacks, we need to develop formalisms, reasoning techniques, and verification methods, to specify systems and protocols, their intended properties, and to guarantee that these intended properties of correctness and security are indeed satisfied.

In Comète we study formal computational frameworks for specifying these systems, theories for defining the desired properties of correctness and security and for reasoning about them, and methods and techniques for proving that a given system satisfies the intended properties.

3. Research Program

3.1. Probability and information theory

Participants: Konstantinos Chatzikokolakis, Catuscia Palamidessi, Marco Romanelli, Anna Pazii.

Much of the research of Comète focuses on security and privacy. In particular, we are interested in the problem of the leakage of secret information through public observables.

Ideally we would like systems to be completely secure, but in practice this goal is often impossible to achieve. Therefore, we need to reason about the amount of information leaked, and the utility that it can have for the adversary, i.e. the probability that the adversary is able to exploit such information.

The recent tendency is to use an information theoretic approach to model the problem and define the leakage in a quantitative way. The idea is to consider the system as an information-theoretic *channel*. The input represents the secret, the output represents the observable, and the correlation between the input and output (*mutual information*) represents the information leakage.

Information theory depends on the notion of entropy as a measure of uncertainty. From the security point of view, this measure corresponds to a particular model of attack and a particular way of estimating the security threat (vulnerability of the secret). Most of the proposals in the literature use Shannon entropy, which is the most established notion of entropy in information theory. We, however, consider also other notions, in particular Rényi min-entropy, which seems to be more appropriate for security in common scenarios like one-try attacks.

3.2. Expressiveness of Concurrent Formalisms

Participants: Catuscia Palamidessi, Frank Valencia.

We study computational models and languages for distributed, probabilistic and mobile systems, with a particular attention to expressiveness issues. We aim at developing criteria to assess the expressive power of a model or formalism in a distributed setting, to compare existing models and formalisms, and to define new ones according to an intended level of expressiveness, also taking into account the issue of (efficient) implementability.

3.3. Concurrent constraint programming

Participants: Frank Valencia, Santiago Quintero.

Concurrent constraint programming (ccp) is a well established process calculus for modeling systems where agents interact by posting and asking information in a store, much like in users interact in *social networks*. This information is represented as first-order logic formulae, called constraints, on the shared variables of the system (e.g., $X > 42$). The most distinctive and appealing feature of ccp is perhaps that it unifies in a single formalism the operational view of processes based upon process calculi with a declarative one based upon first-order logic. It also has an elegant denotational semantics that interprets processes as closure operators (over the set of constraints ordered by entailment). In other words, any ccp process can be seen as an idempotent, increasing, and monotonic function from stores to stores. Consequently, ccp processes can be viewed as: computing agents, formulae in the underlying logic, and closure operators. This allows ccp to benefit from the large body of techniques of process calculi, logic and domain theory.

Our research in ccp develops along the following two lines:

1. **(a)** The study of a bisimulation semantics for ccp. The advantage of bisimulation, over other kinds of semantics, is that it can be efficiently verified.
2. **(b)** The extension of ccp with constructs to capture emergent systems such as those in social networks and cloud computing.

3.4. Model checking

Participants: Konstantinos Chatzikokolakis, Catuscia Palamidessi.

Model checking addresses the problem of establishing whether a given specification satisfies a certain property. We are interested in developing model-checking techniques for verifying concurrent systems of the kind explained above. In particular, we focus on security and privacy, i.e., on the problem of proving that a given system satisfies the intended security or privacy properties. Since the properties we are interested in have a probabilistic nature, we use probabilistic automata to model the protocols. A challenging problem is represented by the fact that the interplay between nondeterminism and probability, which in security presents subtleties that cannot be handled with the traditional notion of a scheduler,

4. Application Domains

4.1. Security and privacy

Participants: Catuscia Palamidessi, Konstantinos Chatzikokolakis, Ehab Elsalamouny, Ali Kassem, Anna Pazii, Marco Romanelli, Natasha Fernandes.

The aim of our research is the specification and verification of protocols used in mobile distributed systems, in particular security protocols. We are especially interested in protocols for *information hiding*.

Information hiding is a generic term which we use here to refer to the problem of preventing the disclosure of information which is supposed to be secret or confidential. The most prominent research areas which are concerned with this problem are those of *secure information flow* and of *privacy*.

Secure information flow refers to the problem of avoiding the so-called *propagation* of secret data due to their processing. It was initially considered as related to software, and the research focussed on type systems and other kind of static analysis to prevent dangerous operations, Nowadays the setting is more general, and a large part of the research effort is directed towards the investigation of probabilistic scenarios and treaths.

Privacy denotes the issue of preventing certain information to become publicly known. It may refer to the protection of *private data* (credit card number, personal info etc.), of the agent's identity (*anonymity*), of the link between information and user (*unlinkability*), of its activities (*unobservability*), and of its *mobility* (*untraceability)*.

The common denominator of this class of problems is that an adversary can try to infer the private information (*secrets*) from the information that he can access (*observables*). The solution is then to obfuscate the link between secrets and observables as much as possible, and often the use randomization, i.e. the introduction of *noise*, can help to achieve this purpose. The system can then be seen as a *noisy channel*, in the information-theoretic sense, between the secrets and the observables.

We intend to explore the rich set of concepts and techniques in the fields of information theory and hypothesis testing to establish the foundations of quantitive information flow and of privacy, and to develop heuristics and methods to improve mechanisms for the protection of secret information. Our approach will be based on the specification of protocols in the probabilistic asynchronous π-calculus, and the application of model-checking to compute the matrices associated to the corresponding channels.

5. Highlights of the Year

5.1. Highlights of the Year

Catuscia Palamidessi has received an European Research Council (ERC) grant for the project HYPATIA.

6. New Software and Platforms

6.1. libqif - A Quantitative Information Flow C++ Toolkit Library

KEYWORDS: Information leakage - Privacy - C++ - Linear optimization

FUNCTIONAL DESCRIPTION: The goal of libqif is to provide an efficient C++ toolkit implementing a variety of techniques and algorithms from the area of quantitative information flow and differential privacy. We plan to implement all techniques produced by Comète in recent years, as well as several ones produced outside the group, giving the ability to privacy researchers to reproduce our results and compare different techniques in a uniform and efficient framework.

Some of these techniques were previously implemented in an ad-hoc fashion, in small, incompatible with each-other, non-maintained and usually inefficient tools, used only for the purposes of a single paper and then abandoned. We aim at reimplementing those – as well as adding several new ones not previously implemented – in a structured, efficient and maintainable manner, providing a tool of great value for future research. Of particular interest is the ability to easily re-run evaluations, experiments and case-studies from all our papers, which will be of great value for comparing new research results in the future.

The library's development continued in 2018 with several new added features. 82 new commits were pushed to the project's git repository during this year. The new functionality was directly applied to the exeperimental results of several publications of the team (QEST'18, Entropy'18, POST'18, CSF'18).

- Contact: Konstantinos Chatzikokolakis
- URL: https://github.com/chatziko/libqif

6.2. F-BLEAU

KEYWORDS: Information leakage - Machine learning - Privacy

FUNCTIONAL DESCRIPTION: F-BLEAU is a tool for estimating the leakage of a system about its secrets in a black-box manner (i.e., by only looking at examples of secret inputs and respective outputs). It considers a generic system as a black-box, taking secret inputs and returning outputs accordingly, and it measures how much the outputs "leak" about the inputs.

F-BLEAU is based on the equivalence between estimating the error of a Machine Learning model of a specific class and the estimation of information leakage.

This code was also used for the experiments of a paper under submission, on the following evaluations: Gowalla, e-passport, and side channel attack to finite field exponentiation.

RELEASE FUNCTIONAL DESCRIPTION: First F-BLEAU release. Supports frequentist and k-NN estimates with several parameters, and it allows stopping according to delta-convergence criteria.

- Contact: Konstantinos Chatzikokolakis
- URL: https://github.com/gchers/fbleau

6.3. Location Guard

KEYWORDS: Privacy - Geolocation - Browser Extensions

SCIENTIFIC DESCRIPTION: The purpose of Location Guard is to implement obfuscation techniques for achieving location privacy, in a an easy and intuitive way that makes them available to the general public. Various modern applications, running either on smartphones or on the web, allow third parties to obtain the user's location. A smartphone application can obtain this information from the operating system using a system call, while web application obtain it from the browser using a JavaScript call.

FUNCTIONAL DESCRIPTION: Websites can ask the browser for your location (via JavaScript). When they do so, the browser first asks your permission, and if you accept, it detects your location (typically by transmitting a list of available wifi access points to a geolocation provider such as Google Location Services, or via GPS if available) and gives it to the website.

Location Guard is a browser extension that intercepts this procedure. The permission dialog appears as usual, and you can still choose to deny. If you give permission, then Location Guard obtains your location and adds "random noise" to it, creating a fake location. Only the fake location is then given to the website.

Location Guard is by now a stable tool with a large user base. No new features were added in 2018, however the tool is still actively maintained, and several issues have beed fixed during this year (new geocoder API, manual installation method for Opera users, etc).

- Participants: Catuscia Palamidessi, Konstantinos Chatzikokolakis, Marco Stronati, Miguel Andrés and Nicolas Bordenabe
- Contact: Konstantinos Chatzikokolakis
- URL: https://github.com/chatziko/location-guard

6.4. dspacenet

Distributed-Spaces Network.

KEYWORDS: Social networks - Distributed programming

FUNCTIONAL DESCRIPTION: DSpaceNet is a tool for social networking based on multi-agent spatial and timed concurrent constraint language.

I - The fundamental structure of DSPaceNet is that of *space*: A space may contain

(1) spatial-mobile-reactive tcc programs, and (2) other spaces.

Furthermore, (3) each space belongs to a given agent. Thus, a space of an agent j within the space of agent i means that agent i allows agent j to use a computation sub-space within its space.

II - The fundamental operation of DSPaceNet is that of *program posting*: In each time unit, agents can post spatial-mobile-reactive tcc programs in the spaces they are allowed to do so (ordinary message posting corresponds to the posting of tell processes). Thus, an agent can for example post a watchdog tcc process to react to messages in their space, e.g. whenever (*happy b*frank*) do tell("thank you!"). More complex mobile programs are also allowed (see below).

The language of programs is a spatial mobile extension of tcc programs:

$P, Q... := tell(c)|whencdoP||nextP|P||Q|unlesscnextP|[P]_i|\uparrow_i P|recX.P$

Computation of timed processes proceeds as in tcc. The spatial construct $[P]_i$ runs P in the space of agent i and the mobile process $\uparrow_i P$, extrudes P from the space of i. By combining space and mobility, arbitrary processes can be moved from one a space into another. For example, one could send a trojan watchdog to another space for spying for a given message and report back to one's space.

III- Constraint systems can be used to specify advance text message deduction, arithmetic deductions, scheduling, etc.

IV - Epistemic Interpretation of spaces can be used to derive whether they are users with conflicting/inconsistent information, or whether a group of agents may be able to deduce certain message.

V - The scheduling of agent requests for program posts, privacy settings, friendship lists are handled by an external interface. For example, one could use type systems to check whether a program complies with privacy settings (for example checking that the a program does not move other program into a space it is not allowed into).

- Partner: Pontificia Universidad Javeriana Cali
- Contact: Frank Valencia
- URL: http://www.dspacenet.com

7. New Results

7.1. Foundations of privacy and quantitative information flow

Privacy and information flow have the common goal of trying to protect sensitive information. Comete focuses in particular on the potential leaks due to inference from data that are public, or anyway available to the adversary. We consider the probabilistic aspects, and we use concepts and tools from information theory.

7.1.1. Black-box Leakage Estimation

In [16] we have considered the problem of measuring how much a system reveals about its secret inputs under the black-box setting. Black-box means that we assume no prior knowledge of the system's internals: the idea is to run the system for choices of secrets and measure its leakage from the respective outputs. Our goal was to estimate the Bayes risk, from which one can derive some of the most popular leakage measures (e.g., min-entropy, additive, and multiplicative leakage). The state-of-the-art method for estimating these leakage measures is the frequentist paradigm, which approximates the system's internals by looking at the frequencies of its inputs and outputs. Unfortunately, this does not scale for systems with large output spaces, where it would require too many input-output examples. Consequently, it also cannot be applied to systems with continuous outputs (e.g., time side channels, network traffic). In [16] we have exploited an analogy between Machine Learning (ML) and black-box leakage estimation to show that the Bayes risk of a system can be estimated by using a class of ML methods: the universally consistent learning rules; these rules can exploit patterns in the input-output examples to improve the estimates' convergence, while retaining formal optimality guarantees. We have focused on a set of them, the nearest neighbor rules; we show that they significantly reduce the number of black-box queries required for a precise estimation whenever nearby outputs tend to be produced by the same secret; furthermore, some of them can tackle systems with continuous outputs. We have illustrated the applicability of these techniques on both synthetic and real-world data, and we compared them with the state-of-the-art tool, leakiEst, which is based on the frequentist approach.

7.1.2. An Axiomatization of Information Flow Measures

Quantitative information flow aims to assess and control the leakage of sensitive information by computer systems. A key insight in this area is that no single leakage measure is appropriate in all operational scenarios; as a result, many leakage measures have been proposed, with many different properties. To clarify this complex situation, in [11] we have studied information leakage axiomatically, showing important dependencies among different axioms. We have also established a completeness result about the g-leakage family, showing that any leakage measure satisfying certain intuitively-reasonable properties can be expressed as a g-leakage.

7.1.3. Comparing systems: max-case refinement orders and application to differential privacy

Quantitative Information Flow (QIF) and Differential Privacy (DP) are both concerned with the protection of sensitive information, but they are rather different approaches. In particular, QIF considers the expected probability of a successful attack, while DP (in both its standard and local versions) is a max-case measure, in the sense that it is compromised by the existence of a possible attack, regardless of its probability. Comparing systems is a fundamental task in these areas: one wishes to guarantee that replacing a system A by a system B is a safe operation, that is the privacy of B is no-worse than that of A. In QIF, a refinement order provides strong such guarantees, while in DP mechanisms are typically compared (wrt privacy) based on the ε privacy parameter that they provide.

In [15] we have explored a variety of refinement orders, inspired by the one of QIF, providing precise guarantees for max-case leakage. We have studied simple structural ways of characterizing them, the relation between them, efficient methods for verifying them and their lattice properties. Moreover, we have applied these orders in the task of comparing DP mechanisms, raising the question of whether the order based on ε provides strong privacy guarantees. We have shown that, while it is often the case for mechanisms of the same "family" (geometric, randomised response, etc.), it rarely holds across different families.

7.1.4. A Logical Characterization of Differential Privacy

Differential privacy (DP) is a formal definition of privacy ensuring that sensitive information relative to individuals cannot be inferred by querying a database. In [12], we have exploited a modeling of this framework via labeled Markov Chains (LMCs) to provide a logical characterization of differential privacy: we have considered a probabilistic variant of the Hennessy-Milner logic and we have defined a syntactical distance on formulae in it measuring their syntactic disparities. Then, we have defined a trace distance on LMCs in terms of the syntactic distance between the sets of formulae satisfied by them. We have proved that such distance corresponds to the level of privacy of the LMCs. Moreover, we have used the distance on formulae to define a real-valued semantics for them, from which we have obtained a logical characterization of weak anonymity: the level of anonymity is measured in terms of the smallest formula distinguishing the considered LMCs. Then, we have focused on bisimulation semantics on nondeterministic probabilistic processes and we have provided a logical characterization of generalized bisimulation metrics, namely those defined via the generalized Kantorovich lifting. Our characterization is based on the notion of mimicking formula of a process and the syntactic distance on formulae, where the former captures the observable behavior of the corresponding process and allows us to characterize bisimilarity. We have shown that the generalized bisimulation distance on processes is equal to the syntactic distance on their mimicking formulae. Moreover, we have used the distance on mimicking formulae to obtain bounds on differential privacy.

7.1.5. Geo-indistinguishability vs Utility in Mobility-based Geographic Datasets

In [17] we have explored the trade-offs between privacy and utility in mobility-based geographic datasets. Our aim was to find out whether it is possible to protect the privacy of the users in a dataset while, at the same time, maintaining intact the utility of the information that it contains. In particular, we have focused on geo-indistinguishability as a privacy-preserving sanitization methodology, and we have evaluated its effects on the utility of the Geolife dataset. We have tested the sanitized dataset in two real world scenarios: 1. Deploying an infrastructure of WiFi hotspots to offload the mobile traffic of users living, working, or commuting in a wide geographic area; 2. Simulating the spreading of a gossip-based epidemic as the outcome of a device-to-device communication protocol. We have shown the extent to which the current geo-indistinguishability techniques trade privacy for utility in real world applications and we focus on their effects at the levels of the population as a whole and of single individuals.

7.1.6. Utility-Preserving Privacy Mechanisms for Counting Queries

Differential privacy(DP) and local differential privacy(LPD) are frameworks to protect sensitive information in data collections. They are both based on obfuscation. In DP the noise is added to the result of queries on the dataset, whereas in LPD the noise is added directly on the individual records, before being collected. The main advantage of LPD with respect to DP is that it does not need to assume a trusted third party. The main disadvantage is that the trade-off between privacy and utility is usually worse than in DP, and typically to retrieve reasonably good statistics from the locally sanitized data it is necessary to have a huge collection of them. In [25], we focus on the problem of estimating counting queries from collections of noisy answers, and we propose a variant of LDP based on the addition of geometric noise. Our main result is that the geometric noise has a better statistical utility than other LPD mechanisms from the literature.

7.1.7. Differential Inference Testing: A Practical Approach to Evaluate Sanitizations of Datasets

In order to protect individuals' privacy, data have to be "well-sanitized" before sharing them, i.e. one has to remove any personal information before sharing data. However, it is not always clear when data shall be deemed well-sanitized. In this paper, we argue that the evaluation of sanitized data should be based on whether the data allows the inference of sensitive information that is specific to an individual, instead of being centered around the concept of re-identification. In [20] we have proposed a framework to evaluate the effectiveness of different sanitization techniques on a given dataset by measuring how much an individual's record from the sanitized dataset influences the inference of his/her own sensitive attribute. Our intent was not to accurately predict any sensitive attribute but rather to measure the impact of a single record on the inference of sensitive information. We have demonstrated our approach by sanitizing two real datasets in different privacy models and evaluate/compare each sanitized dataset in our framework.

7.2. Foundations of Process Calculi

7.2.1. Group Distributed Knowledge.

We introduced spatial constraint systems (scs) as semantic structures for reasoning about spatial and epistemic information in concurrent systems. They have been used to reason about beliefs, lies, and group epistemic behaviour inspired by social networks. They have also been used for proving new results about modal logics and giving semantics to process calculi. In [19] we developed the theory of scs to reason about the distributed information of potentially infinite groups. We characterized the notion of distributed information of a group of agents as the infimum of the set of join-preserving functions that represent the spaces of the agents in the group. We provided an alternative characterization of this notion as the greatest family of join-preserving functions that satisfy certain basic properties. We showed compositionality results for these characterizations and conditions under which information that can be obtained by an infinite group can also be obtained by a finite group. Finally, we provided algorithms that compute the distributive group information of finite groups. Furthermore, in [14] we summarized all the main results we have obtained about scs.

7.2.2. Group Polarization.

Social networks can make their users become more radical and isolated in their own ideological circle causing dangerous splits in society in a phenomenon known as group polarization. In [22] we developed a preliminary model for social networks, and a measure of the level of polarization in these social networks, based on Esteban and Ray's classic measure of polarization for economic situations. Our model includes information about each agent's quantitative strength of belief in a proposition of interest and a representation of the strength of each agent's influence on every other agent. We considered how the model changes over time as agents interact and communicate, and included several different options for belief update, including rational belief update and update taking into account irrational responses such as confirmation bias and the backfire effect. Under various scenarios, we considered the evolution of polarization over time, and the implications of these results for real world social networks.

7.2.3. Lattice Theory.

Structures involving a lattice and join-endomorphisms on it are ubiquitous in computer science. In [28] we studied the cardinality of the set $J(L)$ of all join-endomorphisms of a given finite lattice L. We showed that the cardinality of $J(L)$ is sub-exponential, exponential and super-exponential in the size of the lattice for boolean algebras, linear-orders, and arbitrary lattices, respectively. We also studied the following problem: Given a lattice L of size n and a subset S of $J(L)$ of size m, find the greatest lower bound in $J(L)$ of S. This join-endomorphism has meaningful interpretations in epistemic logic, distributed systems, and Aumann structures. We showed that this problem can be solved with worst-case time complexity in $O(n + mlogn)$ for powerset lattices, $O(mn^2)$ for lattices of sets, and $O(mn + n^3)$ for arbitrary lattices. The complexity is expressed in terms of the basic binary lattice operations performed by the algorithm.

7.2.4. Festschrift Contribution.

In a Festschrift dedicated to Catuscia Palamidessi [26], we presented an article with original solutions to four challenging mathematical puzzles [23]. The first two are concerned with random processes. The first problem can be reduced to computing, for arbitrary large values of n, the expected number of iterations of a program that increases a variable at random between 1 and n until exceeds n. The second problem can be reduced to determining the probability of reaching a given point after visiting all the others in a circular random walk. The other two problems involve finding optimal winning group strategies in guessing games.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. LOST2DNN

Program: DATAIA Call for Research Projects

Project title: Leakage of Sensitive Training Data from Deep Neural Networks

Duration: October 2019 - September 2022

Coordinators: Catuscia Palamidessi, Inria Saclay, EPI Comète and Pablo Piantanida, Centrale Supèlec

Other PI's and partner institutions: Georg Pichler, TU Wien, Austria

Abstract: The overall project goal is to develop a fundamental understanding with experimental validation of the information-leakage of training data from deep learning systems. More specifically, we aim at:

- Developing a compelling case study based on state-of-the-art algorithms to perform model inversion attacks, showcasing the feasibility of uncovering specified sensitive information from a trained software (model) on real data.
- Quantifying information leakage. Based on the uncovered attacks, the amount of sensitive information present in trained software will be measured or quantified. The resulting measure of leakage will serve as a basis for the analysis of attacks and for the development of robust mitigation techniques.
- Mitigating information leakage. Strategies will be explored to avoid the uncovered attacks and minimize the potential information leakage of a trained model.

8.2. National Initiatives

8.2.1. REPAS

Program: ANR Blanc

Project title: Reliable and Privacy-Aware Software Systems via Bisimulation Metrics

Duration: October 2016 - September 2021

Coordinator: Catuscia Palamidessi, Inria Saclay, EPI Comète

Other PI's and partner institutions: Ugo del Lago, Inria Sophia Antipolis (EPI Focus) and University of Bologna (Italy) Vincent Danos, ENS Paris. Filippo Bonchi, ENS Lyon

Abstract: In this project we investigate quantitative notions and tools for proving program correctness and protecting privacy. In particular, we focus on bisimulation metrics, which are the natural extension of bisimulation on quantitative systems. As a key application, we will develop a mechanism to protect the privacy of users when their location traces are collected

8.3. European Initiatives: FP7 & H2020 Projects

8.3.1. HYPATIA

Program: European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.

Project acronym: HYPATIA

Project title: Privacy and Utility Allied

Duration: October 2019 – September 2024

Principal Investigator: Catuscia Palamidessi

Abstract: With the ever-increasing use of internet-connected devices, such as computers, smart grids, IoT appliances and GPS-enabled equipments, personal data are collected in larger and larger amounts, and then stored and manipulated for the most diverse purposes. Undeniably, the big-data technology provides enormous benefits to industry, individuals and society, ranging from improving business strategies and boosting quality of service to enhancing scientific progress. On the other hand, however, the collection and manipulation of personal data raises alarming privacy issues. Not only the experts, but also the population at large are becoming increasingly aware of the risks, due to the repeated cases of violations and leaks that keep hitting the headlines.

The objective of this project is to develop the theoretical foundations, methods and tools to protect the privacy of the individuals while letting their data to be collected and used for statistical purposes. We aim in particular at developing mechanisms that can be applied and controlled directly by the user thus avoiding the need of a trusted party, are robust with respect to combination of information from different sources, and provide an optimal trade-off between privacy and utility.

8.4. International Initiatives

8.4.1. Inria Associate Teams Not Involved in an Inria International Labs

8.4.1.1. LOGIS

Title: Logical and Formal Methods for Information Security

Inria principal investigator: Konstantinos Chatzikokolakis

International Partners:

- Mitsuhiro Okada, Keio University (Japan)
- Yusuke Kawamoto, AIST (Japan)
- Tachio Terauchi, JAIST (Japan)
- Masami Hagiya, University of Tokyo (Japan)

Start year: January 2019 - December 2021.

URL: http://www.lix.polytechnique.fr/~kostas/projects/logis/

Abstract: The project aims at integrating the logical / formal approaches to verify security protocols with (A) complexity theory and (B) information theory. The first direction aims at establishing the foundations of logical verification for security in the computational sense, with the ultimate goal of automatically finding attacks that probabilistic polynomial-time adversaries can carry out on protocols. The second direction aims at developing frameworks and techniques for evaluating and reducing information leakage caused by adaptive attackers.

8.4.2. Inria International Partners

Geoffrey Smith, Florida International University, USA

Carroll Morgan, NICTA , Australia

Annabelle McIver, Maquarie University, Australia

Mario Ferreira Alvim Junior, Assistant Professor, Federal University of Minas Gerais, Brazil

Camilo Rueda, Professor, Universidad Javeriana de Cali, Colombia

Carlos Olarte, Universidade Federal do Rio Grande do Norte, Brazil

Camilo Rocha, Associate Professor, Universidad Javeriana de Cali, Colombia

8.4.3. Participation in Other International Programs

8.4.3.1. CLASSIC

Program: Colciencias - Conv. 712.

Project acronym: CLASSIC.

Project title: Concurrency, Logic and Algebra for Social and Spatial Interactive Computation.

Duration: Oct 2016 - Oct 2019.

URL: http://goo.gl/Gv6Lij

Coordinator: Camilo Rueda, Universidad Javeriana de Cali, Colombia.

Other PI's and partner institutions: Carlos Olarte, Universidade Federal do Rio Grande do Norte, Brazil and Frank Valencia, CNRS-LIX and Inria Saclay.

Abstract: This project will advance the state of the art of domains such as mathematical logic, order theory and concurrency for reasoning about spatial and epistemic behaviour in multi-agent systems..

8.4.3.2. FACTS

Program: ECOS NORD.

Project acronym: FACTS.

Project title: Foundational Approach to Cognition in Today?s Society.

Duration: Jan 1 2019 - Dec 31, 2021.
URL: https://goo.gl/zVhg32
Coordinator: Frank Valencia, Ecole Polytechnique.
Other PI's and partner institutions: Jean-Gabriel Ganascia LIP6, Sorbonne University and Camilo Rueda, Universidad Javeriana de Cali, Colombia.
Abstract: This projects aims at studying the phenomenon of "Group Polarization"; the tendency for a group to learn or acquire beliefs or to make decisions that are more extreme than the initial inclinations of its members.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Yusuke Kawamoto, Researcher, AIST, Japan, AIST, March 2019 and Nov-Dec 2019
Takao Murakami, Researcher, AIST, Japan, AIST, March 2019
Sophia Knight, Assistant Professor, University of Minnesota, USA, May 2019
Carlos Olarte, Assistant Professor, Universidade Federal do Rio Grande do Norte, Brazil. Nov 2019
Camilo Rueda, Professor, Universidad Javeriana de Cali, Colombia. May-July 2019
Mario Ferreira Alvim Junior, Assistant Professor, Federal University of Minas Gerais, Brazil. Nov 2019
Sergio Ramirez, PhD student, Universidad Javeriana de Cali, Colombia. Oct-Dec 2019
Carlos Pinzon, Master student, Universidad Javeriana de Cali, Colombia. Nov 2019

8.5.2. Internships

Sayan Biswas, Master student, Univ. of Bath, UK. From Jun 2019 until Sep 2019
Noemie Fong, Master student, ENS Paris. Jan-Feb 2019
Federica Granese, Univ. Od Rome "La Sapienza", Italy. From Mar 2019 until Jun 2019
Boammani Lompo, ENS Rennes. From May 2019 until Jul 2019

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. Member of the organizing committee

Catuscia Palamidessi is member of:

The Scientific Advisory Board of ANSSI, the French National Cybersecurity Agency. Since 2019.
The Scientific Advisory Board of CISPA, the Helmholtz Center for Information Security. Since 2019.
The Executive Committee of SIGLOG, the ACM Special Interest Group on Logic and Computation. 2014-19.
The Steering Committee of CONCUR, the International Conference in Concurrency Theory. Since 2016.
The Steering Committee of ETAPS, the European Joint Conferences on Theory and Practice of Software. Since 2006.
The Steering Committee of EACSL, the European Association for Computer Science Logics. Since 2015.
The Steering Committee of FORTE, the International Conference on Formal Techniques for Distributed Objects, Components, and Systems. Since 2014.
The IFIP Technical Committee 1 – Foundations of Computer Science. Since 2007.
The IFIP Working Group 1.7 – Theoretical Foundations of Security Analysis and Design. Since 2010.
The IFIP Working Group 1.8 – Concurrency Theory. Since 2005.

Frank D. Valencia is member of:

The steering committee of the International Workshop in Concurrency EXPRESS. Since 2010.

Konstantinos Chatzikokolakis is member of:

The steering committee of the Privacy Enhancing Technologies Symposium. Since 2018.

9.1.2. Scientific events selection committee

9.1.2.1. Chair of conference program committee

Konstantinos Chatzikokolakis:

is serving as PC chair (with Carmela Troncoso as co-chair) of PETS 2020: The 20th Privacy Enhancing Technologies Symposium, July 14 – 18, 2020 Montréal, Canada.

9.1.2.2. Member of conference program committees

Catuscia Palamidessi is/has been a member of the program committees of the following conferences and workshops:

CCS 2020. The ACM Conference on Computer and Communications Security. Orlando, USA, November 9-13 2020.

CSF 2020. The 33rd IEEE Computer Security Foundations Symposium. Boston, MA, USA, June 22-26, 2020.

PETS 2020. The 20th Privacy Enhancing Technologies Symposium. Montréal, Canada, July 14 – 18, 2020.

FORTE 2020. The 40th IFIP International Conference on Formal Techniques for Distributed Objects, Components, and Systems. University of Malta, Valletta, June 15-19, 2020.

FSTTCS 2019. 39th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science. Indian Institute of Technology Bombay, December 11–13, 2019

FACS 2019. The 16th International Conference on Formal Aspects of Component Software. Amsterdam, The Netherlands, 23-25 October 2019.

PETS 2019. The 19th Privacy Enhancing Technologies Symposium. Stockholm, Sweden, July 16 – 20, 2019.

LICS 2019. The Thirty-Fourth Annual ACM/IEEE Symposium on Logic in Computer Science. Vancouver, Canada, 24–27 June 2019.

CSF 2019. The 32nd IEEE Computer Security Foundations Symposium. Hoboken, NJ, USA, June 24-27, 2019.

SAC 2019 (Security track). The 34th ACM/SIGAPP Symposium On Applied Computing. Limassol, Cyprus, 8-12 April 2019.

TML 2020. Towards Trustworthy ML: Rethinking Security and Privacy for ML. Addis Ababa, Ethiopia, April 26, 2020.

EXPRESS/SOS 2020. Combined 27th International Workshop on Expressiveness in Concurrency and 17th Workshop on Structural Operational Semantics. Vienna, Austria, August 31, 2020.

PPAI 2020. The rAAI Workshop on Privacy-Preserving Artificial Intelligence. New York, USA, February 7, 2020.

PPML 2019. Privacy Preserving Machine Learning (ACM CCS 2019 Workshop). London, UK, November 15, 2019.

WPES 2019. Workshop on Privacy in the Electronic Society. London, UK, November 11, 2019.

APVP 2019. Atelier sur la Protection de la Vie Privée. Cap Hornu, France, July 9-11, 2019.

WIL 2019. 3rd Women in Logic Workshop. Cap Hornu, France, Vancouver, Canada, June 23, 2019.

Frank D. Valencia is/has been a member of the program committees of the following conferences and workshops:

CP-DP-19: Doctoral Program of the 25th International Conference on Principles and Practice of Constraint Programming. Stamford, CT, USA, Sep 30 - Oct 4, 2019.

CONCUR 2019: The 30th International Conference on Concurrency Theory. Amsterdam, Netherlands, August 26-31, 2019.

AAMAS 2019: International Conference on Autonomous Agents and Multiagent Systems. Montreal, Canada, 13th-17th of May 2019.

9.1.3. Journals

9.1.3.1. Member of the editorial board

Catuscia Palamidessi is:

(2019-) Member of the Editorial Board of the Journal of Computer Security IOS Press. Since 2019.

Member of the Editorial Board of the Proceedings on Privacy Enhancing Technologies (PoPETs), published by De Gruyter. Since 2017.

Member of the Editorial Board of Mathematical Structures in Computer Science, published by the Cambridge University Press. Since 2006.

Member of the Editorial Board of Acta Informatica, published by Springer. Since 2015.

Member of the Editorial Board of the Electronic Notes of Theoretical Computer Science, published by Elsevier Science. Since 2000.

Member of the Editorial Board of LIPIcs: Leibniz International Proceedings in Informatics, Schloss Dagstuhl–Leibniz Center for Informatics. Since 2014.

Konstantinos Chatzikokolakis is:

Editorial board member of the Proceedings on Privacy Enhancing Technologies (PoPETs), a scholarly journal for timely research papers on privacy.

9.1.3.2. Reviewing

The members of the team regularly review papers for international journals, conferences and workshops.

9.1.4. Other Editorial Activities

Catuscia Palamidessi is/has been:

Co-editor (with Anca Muscholl and Anuj Dawar) of the special issue of Logical Methods in Computer Science dedicated to selected papers of ICALP 2017.

Co-editor (with Alexandra Silva and Natarajan Shankar) of the special issue of Logical Methods in Computer Science dedicated to selected papers of LICS 2015 and LICS 2016.

Frank D. Valencia has been:

Co-editor of the special issue on Mathematical Structures in Computer Science dedicated to the best papers from the 12th International Colloquium on Theoretical Aspects of Computing.

9.1.5. Participation in other committees

Catuscia Palamidessi has been serving in the following committees:

External Member of the committee for the promotion to full professor of Prof. Kévin Huguenin. HEC Lausanne, Switzerland.

Member of the committee for associate professor positions in the Datalogi Dept., Aalborg Univ., Denmark. 2019.

Member of the panel for the Research Evaluation for Development 2019 (RED19) of the Department of Computer Science and Engineering at the University of Gothenburg, Sweden.

Chair of the Nominating Committee for the 2019 renewal of the office holders of SIGLOG, the ACM Special Interest Group on Logic and Computation.

Member of the committee for the Alonzo Church Award for Outstanding Contributions to Logic and Computation. From 2015. In 2018 Palamidessi is the president of this committee.

Reviewer for the projects proposal for the program PRIN, sponsored by the Italian MIUR ("Ministero dell'Istruzione, dell'Università e della Ricerca"). Since 2005.

Member of the EAPLS PhD Award Committee. From 2010.

9.1.6. Invited talks

Catuscia Palamidessi has given invited talks at the following conferences and workshops:

Journées Nationales du GDR Sécurité Informatique. Paris, France, June 2019.

MFPS XXXV, special session on Probabilistic Programming. London, UK, June 2019.

AI & Society: From principles to practice. CIFAR-UKRI-CNRS workshop. London, UK, June 2019.

5th France-Japan Cybersecurity workshop}. Kyoto, Japan, April 2019.

Frank Valencia has given the following invited talk:

- EXPRESS/SOS 2019. Combined 26th International Workshop on Expressiveness in Concurrency and 16th Workshop on Structural Operational Semantics. Amsterdam, Netherlands, August 26, 2019.

9.1.7. Service

Catuscia Palamidessi has served as:

Member of the committee for the assignment of the Inria International Chairs. From 2017.

Member of the Commission Scientifique du Centre de Recherche Inria Saclay. From 2018.

Member of the hiring committee for Maitre de Conference, Ecole Polytechnique, 2019.

Frank Valencia has served as:

Directeur adjoint de l'UMR 7161, le Laboratoire d'Informatique de l'Ecole Polytechnique (LIX). May 2016 - Nov 2019.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master : Frank D. Valencia has been teaching the undergraduate course "Computability", 45 hours, at the Pontificia Universidad Javeriana de Cali, Colombia. July 27 - Nov 1, 2019.

Master : Frank D. Valencia has been teaching the masters course "Foundations of Computer Science", 45 hours, at the Pontificia Universidad Javeriana de Cali, Colombia. Jan 27 - Jun 1, 2019

Master : Catuscia Palamidessi has been teaching the masters course on "Foundations of Privacy", 24 hours, at the MPRI, Sept-Nov 2019.

9.2.2. Supervision

PhD in progress (2019-) Federica Granese. Co-supervised Catuscia Palamidessi and Daniele Gorla. Thesis subject: Security in Machine Learning.

PhD in progress (2019-) Ganesh Del Grosso. Co-supervised by Catuscia Palamidessi and Pablo Piantanida. Thesis Subject: Privacy in Machine Learning.

PhD in progress (2018-) Natasha Fernandez. Co-supervised Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Annabelle McIver. Thesis subject: Privacy Protection Methods for Textual Documents.

PhD in progress (2018-) Santiago Quintero. Co-supervised by Frank Valencia and Catuscia Palamidessi. Thesis Subject: Foundations of Group Polarization.

PhD in progress (2017-) Marco Romanelli. Co-supervised by Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Moreno Falaschi (University of Siena, Italy). Thesis subject: Application of Information Flow to feature selection in machine learning.

PhD in progress (2017-) Anna Pazii. Co-supervised by Konstantinos Chatzikokolakis and Catuscia Palamidessi. Thesis subject: Local Differential Privacy.

PhD in progress (2017-) Sergio Ramirez. Co-supervised by Frank Valencia and Camilo Rueda, Universidad Javeriana Cali. Thesis subject: Quantitive Spatial Constraint Systems.

9.2.3. Juries

Catuscia Palamidessi has been reviewer and member of the board at the PhD defense for the thesis of the following PhD student:

Mohamed Maouche (INSA, Lyon). PhD thesis reviewer and member of the committee board at the PhD defense. Title of the thesis: *Protection against Re-identification Attacks in Location Privacy*. Defended in November 2019.

Raphaëlle Crubrillé (IRIF, Université Paris Diderot). Member of the committee board at the PhD defense. Title of the thesis: *Distances comportementales pour les programmes probabilistes d'ordre supérieur*. Defended in June 2019.

Vittoria Nardone (University of Sannio, Italy). PhD thesis reviewer. Title of the thesis: *Formal Methods for Android Applications*. Supervised by Antonella Santone. Defended in January 2019.

9.2.4. Other didactical duties

Catuscia Palamidessi has been:

External member of the scientific council for the PhD in Computer Science at the University of Pisa, Italy. Since 2012.

9.3. Popularization

9.3.1. Education

Konstantinos Chatzikokolakis and Catuscia Palamidessi have designed, and coordinate, a course on the Foundations of Privacy at the MPRI, the Master Parisien pour la Recherche en Informatique. University of Paris VII. A.Y. Since 2015.
Catuscia Palamidessi has been:

- Invited speaker at PLMW@POPL 2019, the Programming Logic Mentoring Workshop 2019 (affiliated to POPL 2019). This workshop aims at encouraging graduate students and senior undergraduate students to pursue careers in programming language research, and at educating them on the research career.
- A participant in the round table on fairness, Interpretability, and Privacy in AI at the CIFAR-UKRI-CNRS workshop. London, June 2019.

Frank Valencia has:

- Welcomed the students of bachelor program of École Polytechnique to Inria center. Sept 12, 2019.
- Welcomed visitors from ACOFI, the Colombian Association of Faculties of Engineering to Inria center. April 8, 2019.

9.3.2. Interventions

Catuscia Palamidessi has given an invited talk at:

- Safety and AI. DATAIA Workshop. Palaiseau, France, September 2019.

10. Bibliography

Major publications by the team in recent years

[1] M. S. ALVIM, M. E. ANDRÉS, K. CHATZIKOKOLAKIS, P. DEGANO, C. PALAMIDESSI.*On the information leakage of differentially-private mechanisms*, in "Journal of Computer Security", 2015, vol. 23, n^{o} 4, p. 427-469 [*DOI :* 10.3233/JCS-150528], https://hal.inria.fr/hal-00940425

[2] M. S. ALVIM, K. CHATZIKOKOLAKIS, A. MCIVER, C. MORGAN, C. PALAMIDESSI, G. SMITH.*Additive and multiplicative notions of leakage, and their capacities*, in "27th Computer Security Foundations Symposium (CSF 2014)", Vienna, Austria, IEEE, July 2014, p. 308–322 [*DOI :* 10.1109/CSF.2014.29], https://hal.inria.fr/hal-00989462

[3] M. S. ALVIM, K. CHATZIKOKOLAKIS, A. MCIVER, C. MORGAN, C. PALAMIDESSI, G. SMITH.*An Axiomatization of Information Flow Measures*, in "Theoretical Computer Science", 2019, vol. 777, p. 32-54 [*DOI :* 10.1016/J.TCS.2018.10.016], https://hal.archives-ouvertes.fr/hal-01995712

[4] M. S. ALVIM, K. CHATZIKOKOLAKIS, C. PALAMIDESSI, G. SMITH.*Measuring Information Leakage using Generalized Gain Functions*, in "Computer Security Foundations", Cambridge MA, United States, IEEE, 2012, p. 265-279 [*DOI :* 10.1109/CSF.2012.26], http://hal.inria.fr/hal-00734044

[5] M. E. ANDRÉS, N. E. BORDENABE, K. CHATZIKOKOLAKIS, C. PALAMIDESSI.*Geo-Indistinguishability: Differential Privacy for Location-Based Systems*, in "20th ACM Conference on Computer and Communications Security", Berlin, Allemagne, ACM Press, 2013, p. 901-914, DGA, Inria large scale initiative CAPPRIS [*DOI :* 10.1145/2508859.2516735], http://hal.inria.fr/hal-00766821

[6] N. E. BORDENABE, K. CHATZIKOKOLAKIS, C. PALAMIDESSI.*Optimal Geo-Indistinguishable Mechanisms for Location Privacy*, in "CCS - 21st ACM Conference on Computer and Communications Security", Scottsdale, Arizona, United States, G.-J. AHN, M. YUNG, N. LI (editors), Proceedings of the 21st ACM Conference on Computer and Communications Security, ACM, November 2014, p. 251-262 [*DOI :* 10.1145/2660267.2660345], https://hal.inria.fr/hal-00950479

[7] G. CHERUBIN, K. CHATZIKOKOLAKIS, C. PALAMIDESSI.*F-BLEAU: Fast Black-Box Leakage Estimation*, in "Proceedings of the 40th IEEE Symposium on Security and Privacy (SP)", San Francisco, United States, IEEE, May 2019, p. 835-852 [*DOI :* 10.1109/SP.2019.00073], https://hal.archives-ouvertes.fr/hal-02422945

[8] M. GUZMÁN, S. KNIGHT, S. QUINTERO, S. RAMÍREZ, C. RUEDA, F. D. VALENCIA.*Reasoning about Distributed Knowledge of Groups with Infinitely Many Agents*, in "CONCUR 2019 - 30th International Conference on Concurrency Theory", Amsterdam, Netherlands, W. FOKKINK, R. VAN GLABBEEK (editors), August 2019, vol. 140, p. 29:1–29:15 [*DOI :* 10.4230/LIPICS.CONCUR.2019.29], https://hal.archives-ouvertes.fr/hal-02172415

[9] M. GUZMÁN, S. HAAR, S. PERCHY, C. RUEDA, F. D. VALENCIA.*Belief, Knowledge, Lies and Other Utterances in an Algebra for Space and Extrusion*, in "Journal of Logical and Algebraic Methods in Programming", September 2016 [*DOI :* 10.1016/J.JLAMP.2016.09.001], https://hal.inria.fr/hal-01257113

[10] S. KNIGHT, C. PALAMIDESSI, P. PANANGADEN, F. D. VALENCIA.*Spatial and Epistemic Modalities in Constraint-Based Process Calculi*, in "CONCUR 2012 - Concurrency Theory - 23rd International Confer-

ence, CONCUR 2012", Newcastle upon Tyne, United Kingdom, September 2012, vol. 7454, p. 317-332 [*DOI :* 10.1007/978-3-642-32940-1], http://hal.inria.fr/hal-00761116

Publications of the year

Articles in International Peer-Reviewed Journal

[11] M. S. ALVIM, K. CHATZIKOKOLAKIS, C. MORGAN, C. PALAMIDESSI, G. SMITH, A. MCIVER.*An Axiomatization of Information Flow Measures*, in "Theoretical Computer Science", 2019, vol. 777, p. 32-54 [*DOI :* 10.1016/J.TCS.2018.10.016], https://hal.archives-ouvertes.fr/hal-01995712

[12] V. CASTIGLIONI, K. CHATZIKOKOLAKIS, C. PALAMIDESSI.*A Logical Characterization of Differential Privacy*, in "Science of Computer Programming", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02423048

[13] M. FALASCHI, M. GABBRIELLI, C. OLARTE, C. PALAMIDESSI.*Dynamic slicing for Concurrent Constraint Languages*, in "Fundamenta Informaticae", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02423973

[14] F. D. VALENCIA.*Semantic Structures for Spatially-Distributed Multi-Agent Systems*, in "Electronic Proceedings in Theoretical Computer Science", August 2019, vol. 300, p. 39-53 [*DOI :* 10.4204/EPTCS.300.3], https://hal.archives-ouvertes.fr/hal-02410770

International Conferences with Proceedings

[15] K. CHATZIKOKOLAKIS, N. FERNANDES, C. PALAMIDESSI.*Comparing systems: max-case refinement orders and application to differential privacy*, in "Proceedings of the 32nd IEEE Computer Security Foundations Symposium", Hoboken, United States, 2019, p. 442–457 [*DOI :* 10.1109/CSF.2019.00037], https://hal.archives-ouvertes.fr/hal-02126848

[16] G. CHERUBIN, K. CHATZIKOKOLAKIS, C. PALAMIDESSI.*F-BLEAU: Fast Black-Box Leakage Estimation*, in "Proceedings of the 40th IEEE Symposium on Security and Privacy (SP)", San Francisco, United States, IEEE, May 2019, p. 835-852 [*DOI :* 10.1109/SP.2019.00073], https://hal.archives-ouvertes.fr/hal-02422945

[17] A. DI LUZIO, A. C. VIANA, K. CHATZIKOKOLAKIS, G. DIKOV, C. PALAMIDESSI, J. STEFA.*Catch Me If You Can: How Geo-indistinguishability Affects Utility in Mobility-based Geographic Datasets*, in "Proceedings of the 3rd ACM SIGSPATIAL International Workshop", Chicago, United States, ACM Press, 2019, p. 1-10 [*DOI :* 10.1145/3356994.3365498], https://hal.archives-ouvertes.fr/hal-02423337

[18] D. GORLA, F. GRANESE, C. PALAMIDESSI.*Enhanced Models for Privacy and Utility in Continuous-Time Diffusion Networks*, in "Proceedings of the 16th International Colloquium on Theoretical Aspects of Computing (ICTAC)", Hammamet, Tunisia, R. M. HIERONS, M. MOSBAH (editors), Lecture Notes in Computer Science, Springer, October 2019, vol. 11884, p. 313-331 [*DOI :* 10.1007/978-3-030-32505-3_18], https://hal.inria.fr/hal-02424329

[19] M. GUZMÁN, S. KNIGHT, S. QUINTERO, S. RAMÍREZ, C. RUEDA, F. D. VALENCIA.*Reasoning about Distributed Knowledge of Groups with Infinitely Many Agents*, in "CONCUR 2019 - 30th International Conference on Concurrency Theory", Amsterdam, Netherlands, W. FOKKINK, R. VAN GLABBEEK (editors), August 2019, vol. 29, p. 1-29 [*DOI :* 10.4230/LIPICS.CONCUR.2019.29], https://hal.archives-ouvertes.fr/hal-02172415

[20] A. KASSEM, G. ACS, C. CASTELLUCCIA, C. PALAMIDESSI.*Differential Inference Testing: A Practical Approach to Evaluate Sanitizations of Datasets*, in "Proceedings of the 2019 IEEE Security and Privacy Workshops (SPW)", San Francisco, United States, IEEE, May 2019, p. 72-79 [*DOI :* 10.1109/SPW.2019.00024], https://hal.archives-ouvertes.fr/hal-02422992

Scientific Books (or Scientific Book chapters)

[21] M. S. ALVIM, K. CHATZIKOKOLAKIS, A. MCIVER, C. MORGAN, C. PALAMIDESSI, G. SMITH.*The Science of Quantitative Information Flow*, Springer, 2019, forthcoming, https://hal.inria.fr/hal-01971490

[22] M. S. ALVIM, S. KNIGHT, F. D. VALENCIA.*Toward a Formal Model for Group Polarization in Social Networks*, in "The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy - Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday", M. S. ALVIM, K. CHATZIKOKOLAKIS, C. OLARTE, F. D. VALENCIA (editors), Lecture Notes in Computer Science, Springer, November 2019, vol. 11760, p. 419-441 [*DOI :* 10.1007/978-3-030-31175-9_24], https://hal.archives-ouvertes.fr/hal-02410747

[23] N. ARISTIZABAL, C. PINZÓN, C. RUEDA, F. D. VALENCIA.*Make Puzzles Great Again*, in "The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy - Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday", M. S. ALVIM, K. CHATZIKOKOLAKIS, C. OLARTE, F. D. VALENCIA (editors), Lecture Notes in Computer Science, Springer, November 2019, vol. 11760, p. 442-459 [*DOI :* 10.1007/978-3-030-31175-9_25], https://hal.archives-ouvertes.fr/hal-02410767

[24] K. CHATZIKOKOLAKIS, G. SMITH.*Refinement Metrics for Quantitative Information Flow*, in "The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy. Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday", M. S. ALVIM, K. CHATZIKOKOLAKIS, C. OLARTE, F. D. VALENCIA (editors), Lecture Notes in Computer Science, Springer, November 2019, vol. 11760, p. 397-416 [*DOI :* 10.1007/978-3-030-31175-9_23], https://hal.inria.fr/hal-02350777

[25] N. FERNANDES, K. LEFKI, C. PALAMIDESSI.*Utility-Preserving Privacy Mechanisms for Counting Queries*, in "Models, Languages and Tools for Concurrent and Distributed Programming", M. BOREALE, F. CORRADINI, M. LORETI, R. PUGLIESE (editors), Lecture Notes in Computer Science, Springer, July 2019, vol. 11665, p. 487-495, https://arxiv.org/abs/1906.12147 , https://hal.archives-ouvertes.fr/hal-02169218

Books or Proceedings Editing

[26] M. S. ALVIM, K. CHATZIKOKOLAKIS, C. OLARTE, F. D. VALENCIA (editors). *The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy - Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday*, Lecture Notes in Computer Science, Springer, November 2019, vol. 11760 [*DOI :* 10.1007/978-3-030-31175-9], https://hal.archives-ouvertes.fr/hal-02411252

[27] A. DAWAR, A. MUSCHOLL, C. PALAMIDESSI (editors). *Selected Papers of the 44th International Colloquium on Automata, Languages and Programming (ICALP 2017)*, Logical Methods in Computer Science Association, 2019, forthcoming, https://hal.inria.fr/hal-01997414

Research Reports

[28] S. QUINTERO, S. RAMÍREZ, C. RUEDA, F. D. VALENCIA.*Counting and Computing Join-Endomorphisms in Lattices*, LIX, Ecole polytechnique ; Inria Saclay - Ile-de-France, December 2019, Conditionally Accepted at RAMICS 2020, https://hal.archives-ouvertes.fr/hal-02422624

Project-Team COMMANDS

Control, Optimization, Models, Methods and Applications for Nonlinear Dynamical Systems

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization and control of dynamic systems

Table of contents

Project-Team COMMANDS

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

A6.2.1. - Numerical analysis of PDE and ODE
A6.2.6. - Optimization
A6.2.7. - High performance computing
A6.3.2. - Data assimilation
A6.4.1. - Deterministic control
A6.4.2. - Stochastic control

Other Research Topics and Application Domains:

B4.4. - Energy delivery
B4.4.1. - Smart grids
B7.1.2. - Road traffic
B7.1.3. - Air traffic
B7.2.1. - Smart vehicles

1. Team, Visitors, External Collaborators

Research Scientists

Joseph Frederic Bonnans [Team leader, Inria, Senior Researcher, HDR]
Laurent Pfeiffer [Inria, Researcher, from Oct 2019]

External Collaborators

Axel Kröner [U. Humboldt, until Mar 2019]
Laurent Pfeiffer [U. Graz, from Jun 2019 until Sep 2019]

Technical Staff

Hélèna Kutniak [Inria, Engineer, until Jan 2019]

PhD Students

Guillaume Bonnet [Univ Paris-Sud, PhD Student]
Pierre Lavigne [École polytechnique, PhD Student]
Arthur Le Rhun [IFPEN, PhD Student]

Post-Doctoral Fellows

Justina Gianatti [Inria, Post-Doctoral Fellow, until Apr 2019]
Saeed Hadikhanloo [Inria, Post-Doctoral Fellow, until Jan 2019]
Davin Glen Lunz [Inria, Post-Doctoral Fellow, from Nov 2019]

Administrative Assistants

Ines Dumontier [Inria, Administrative Assistant, until May 2019]
Hanadi Dib [Inria, Administrative Assistant,since May 2019]

2. Overall Objectives

2.1. Scientific directions

Commands is a team devoted to dynamic optimization, both for deterministic and stochastic systems. This includes the following approaches: trajectory optimization, deterministic and stochastic optimal control, stochastic programming, dynamic programming and Hamilton-Jacobi-Bellman equation.

Our aim is to derive new and powerful algorithms for solving numerically these problems, with applications in several industrial fields. While the numerical aspects are the core of our approach it happens that the study of convergence of these algorithms and the verification of their well-posedness and accuracy raises interesting and difficult theoretical questions, such as, for trajectory optimization: qualification conditions and second-order optimality condition, well-posedness of the shooting algorithm, estimates for discretization errors; for the Hamilton-Jacobi-Bellman approach: accuracy estimates, strong uniqueness principles when state constraints are present, for stochastic programming problems: sensitivity analysis.

2.2. Industrial impact

For many years the team members have been deeply involved in various industrial applications, often in the framework of PhD theses. The Commands team itself has dealt since its foundation in 2009 with several types of applications:

- Space vehicle trajectories, in collaboration with CNES, the French space agency.
- Aeronautics, in collaboration with the startup Safety Line.
- Production, management, storage and trading of energy resources, in collaboration with Edf, ex-Gdf and Total.
- Energy management for hybrid vehicles, in collaboration with Renault and Ifpen.

We give more details in the Bilateral contracts section.

3. Research Program

3.1. Historical aspects

The roots of deterministic optimal control are the "classical" theory of the calculus of variations, illustrated by the work of Newton, Bernoulli, Euler, and Lagrange (whose famous multipliers were introduced in [24]), with improvements due to the "Chicago school", Bliss [16] during the first part of the 20th century, and by the notion of relaxed problem and generalized solution (Young [29]).

Trajectory optimization really started with the spectacular achievement done by Pontryagin's group [28] during the fifties, by stating, for general optimal control problems, nonlocal optimality conditions generalizing those of Weierstrass. This motivated the application to many industrial problems (see the classical books by Bryson and Ho [20], Leitmann [26], Lee and Markus [25], Ioffe and Tihomirov [23]).

Dynamic programming was introduced and systematically studied by R. Bellman during the fifties. The HJB equation, whose solution is the value function of the (parameterized) optimal control problem, is a variant of the classical Hamilton-Jacobi equation of mechanics for the case of dynamics parameterized by a control variable. It may be viewed as a differential form of the dynamic programming principle. This nonlinear first-order PDE appears to be well-posed in the framework of *viscosity solutions* introduced by Crandall and Lions [21]. The theoretical contributions in this direction did not cease growing, see the books by Barles [14] and Bardi and Capuzzo-Dolcetta [13].

3.2. Trajectory optimization

The so-called *direct methods* consist in an optimization of the trajectory, after having discretized time, by a nonlinear programming solver that possibly takes into account the dynamic structure. So the two main problems are the choice of the discretization and the nonlinear programming algorithm. A third problem is the possibility of refinement of the discretization once after solving on a coarser grid.

In the *full discretization approach*, general Runge-Kutta schemes with different values of control for each inner step are used. This allows to obtain and control high orders of precision, see Hager [22], Bonnans [17]. In the *indirect* approach, the control is eliminated thanks to Pontryagin's maximum principle. One has then to solve the two-points boundary value problem (with differential variables state and costate) by a single or multiple shooting method. The questions are here the choice of a discretization scheme for the integration of the boundary value problem, of a (possibly globalized) Newton type algorithm for solving the resulting finite dimensional problem in IR^n (n is the number of state variables), and a methodology for finding an initial point.

3.3. Hamilton-Jacobi-Bellman approach

This approach consists in calculating the value function associated with the optimal control problem, and then synthesizing the feedback control and the optimal trajectory using Pontryagin's principle. The method has the great particular advantage of reaching directly the global optimum, which can be very interesting when the problem is not convex.

Optimal stochastic control problems occur when the dynamical system is uncertain. A decision typically has to be taken at each time, while realizations of future events are unknown (but some information is given on their distribution of probabilities). In particular, problems of economic nature deal with large uncertainties (on prices, production and demand). Specific examples are the portfolio selection problems in a market with risky and non-risky assets, super-replication with uncertain volatility, management of power resources (dams, gas). Air traffic control is another example of such problems.

For solving stochastic control problems, we studied the so-called Generalized Finite Differences (GFD), that allow to choose at any node, the stencil approximating the diffusion matrix up to a certain threshold [19]. Determining the stencil and the associated coefficients boils down to a quadratic program to be solved at each point of the grid, and for each control. This is definitely expensive, with the exception of special structures where the coefficients can be computed at low cost. For two dimensional systems, we designed a (very) fast algorithm for computing the coefficients of the GFD scheme, based on the Stern-Brocot tree [18].

4. Application Domains

4.1. Energy management for hybrid vehicles

In collaboraton with Ifpen and in the framework of A. Le Rhun's thesis, we have developped a methodology for the optimal energy management for hybrid vehicles, based on a statistical analysis of the traffic. See [12], [12], [7].

4.2. Biological cells culture

In collaboration with the Inbio team (Inst. Pasteur and Inria) we started to study the optimization of protein production based on cell culture.

5. Highlights of the Year

5.1. Highlights of the Year

We have now a strong involvement in the study of mean-field games (MFG) and their application to distributed energy production problems. In the paper [3] we study MFG equilibria with coupling of the agents through a price function (see more in the 'New Results' section). In the framework of the PhD of Pierre Lavigne we currently study discrete-time models with risk-averse agents. Both directions take advantange of the recent recruitment of Laurent Pfeiffer as "chargé de recherche", and of a starting collaboration with Jameson Graber (Baylor University, Texas).

6. New Software and Platforms

6.1. BOCOP

Boite à Outils pour le Contrôle OPtimal

KEYWORDS: Dynamic Optimization - Identification - Biology - Numerical optimization - Energy management - Transportation

FUNCTIONAL DESCRIPTION: Bocop is an open-source toolbox for solving optimal control problems, with collaborations with industrial and academic partners. Optimal control (optimization of dynamical systems governed by differential equations) has numerous applications in transportation, energy, process optimization, energy and biology. Bocop includes a module for parameter identification and a graphical interface, and runs under Linux / Windows / Mac.

RELEASE FUNCTIONAL DESCRIPTION: Handling of delay systems Alternate automatic differentiation tool: CppAD Update for CMake and MinGW (windows version)

- Participants: Benjamin Heymann, Virgile Andréani, Jinyan Liu, Joseph Frédéric Bonnans and Pierre Martinon
- Contact: Pierre Martinon
- URL: http://bocop.org

6.2. Bocop HJB

KEYWORDS: Optimal control - Stochastic optimization - Global optimization

FUNCTIONAL DESCRIPTION: Toolbox for stochastic or deterministic optimal control, dynamic programming / HJB approach.

RELEASE FUNCTIONAL DESCRIPTION: User interface State jumps for switched systems Explicit handling of final conditions Computation of state probability density (fiste step to mean field games)

- Participants: Benjamin Heymann, Jinyan Liu, Joseph Frédéric Bonnans and Pierre Martinon
- Contact: Joseph Frédéric Bonnans
- URL: http://bocop.org

6.3. Bocop Avion

KEYWORDS: Optimization - Aeronautics

FUNCTIONAL DESCRIPTION: Optimize the climb speeds and associated fuel consumption for the flight planning of civil airplanes.

NEWS OF THE YEAR: Improved atmosphere model 2D interpolations for temperature and wind data

- Participants: Gregorutti Baptiste, Cindie Andrieu, Anamaria Lupu, Joseph Frédéric Bonnans, Karim Tekkal, Pierre Jouniaux and Pierre Martinon
- Partner: Safety Line
- Contact: Pierre Martinon
- URL: http://www.safety-line.fr

6.4. Bocop HJB Avion

KEYWORDS: Optimization - Aeronautics

FUNCTIONAL DESCRIPTION: Optimize the climb and cruising trajectory of flight by a HJB approach.

News Of The Year: First demonstrator for cruise flight deployed at Safety Line

- Participants: Pierre Martinon, Joseph Frédéric Bonnans, Jinyan Liu, Gregorutti Baptiste and Ana-maria Lupu
- Partner: Safety Line
- Contact: Pierre Martinon
- URL: http://www.safety-line.fr

7. New Results

7.1. Stochastic control and HJB equations

7.1.1. Monotone and second order consistent schemes for the Pucci and Monge-Ampere equations

In [9] we introduce a new strategy for the design of second-order accurate discretiza-tions of non-linear second order operators of Bellman type, which preserves degenerate ellipticity. The approach relies on Selling's formula, a tool from lattice geometry, and is applied to the Pucci and Monge-Ampere equations, discretized on a two dimensional cartesian grid. In the case of the Monge-Ampere equation, our work is related to both the stable formulation and the second order accurate scheme. Numerical experiments illustrate the robustness and the accuracy of the method.

7.1.2. Mean-field games of control

In [3], an existence result for a class of mean field games of controls is provided. In the considered model, the cost functional to be minimized by each agent involves a price depending at a given time on the controls of all agents and a congestion term. The existence of a classical solution is demonstrated with the Leray–Schauder theorem; the proof relies in particular on a priori bounds for the solution, which are obtained with the help of a potential formulation of the problem.

7.2. Optimal control of PDEs

7.2.1. Optimal Control of an Age-Structured System with State Constraints

In [10] we study an optimal control problem with state constraints where the state is given by an age-structured, abstract parabolic differential equation. We prove the existence and uniqueness of solution for the state equation and provide first and second parabolic estimates. We analyze the differentiability of the cost function and, based on the general theory of Lagrange multipliers, we give a first order optimality condition. We also define and analyze the regularity of the costate. Finally, we present a pregnancy model, where two coupled age-structured equations are involved, and we apply the obtained results to this case.

7.2.2. Feedback laws

The articles [4], [5], [6], co-written by L. Pfeiffer in the framework of his former position at the University of Graz, deal with the computation of feedback laws for stabilization problems of PDE systems. These problems are formulated as infinite-horizon optimal control problems.

In [5], we prove that the value function associated with bilinear stabilization problems (including some control problems of the Fokker-Planck equation) can be expanded as a Taylor expansion, where the second-order term is the solution to an algebraic Riccati equation and where the terms of order three and more are solutions to well-posed linear equations. These equations are obtained by successive differentiation of the HJB equation. A polynomial feedback law can be deduced from the Taylor approximation and its efficiency is analyzed. This approach generalizes the classical LQR-stabilization method.

In [4], we apply the methodology previously described to a stabilization problem of the 2D Navier-Stokes equation. Numerical results are provided.

In [6], we analyze an implementation of the Receding-Horizon Control method utilizing the Taylor expansion of the value function as a terminal cost. More precisely, we show that the method converges at an exponential rate with respect to the prediction horizon and the degree of the Taylor approximation.

7.3. Energy management for hybrid vehicles

7.3.1. A stochastic data-based traffic model applied to vehicles energy consumption estimation

In [7], a new approach to estimate traffic energy consumption via traffic data aggregation in (speed,acceleration) probability distributions is proposed. The aggregation is done on each segment composing the road network. In order to reduce data occupancy, clustering techniques are used to obtain meaningful classes of traffic conditions. Different times of the day with similar speed patterns and traffic behavior are thus grouped together in a single cluster. Different energy consumption models based on the aggregated data are proposed to estimate the energy consumption of the vehicles in the road network. For validation purposes, a microscopic traffic simulator is used to generate the data and compare the estimated energy consumption to the reference one. A thorough sensitivity analysis with respect to the parameters of the proposed method (i.e. number of clusters, size of the distributions support, etc.) is also conducted in simulation. Finally, a real-life scenario using floating car data is analyzed to evaluate the applicability and the robustness of the proposed method.

7.3.2. A bi-level energy management strategy for HEVs under probabilistic traffic conditions

In [11], we propose a new approach to optimize the consumption of a hybrid electric vehicle taking into account the traffic conditions. The method is based on a bi-level decomposition in order to make the implementation suitable for online use. The offline lower level computes cost maps thanks to a stochastic optimization that considers the influence of traffic, in terms of speed/acceleration probability distributions. At the online upper level, a deterministic optimization computes the ideal state of charge at the end of each road segment, using the computed cost maps. Since the high computational cost due to the uncertainty of traffic conditions has been managed at the lower level, the upper level is fast enough to be used online in the vehicle. Errors due to discretization and computation in the proposed algorithm have been studied. Finally, we present numerical simulations using actual traffic data, and compare the proposed bi-level method to a deterministic optimization with perfect information about traffic conditions. The solutions show a reasonable over-consumption compared with deterministic optimization, and manageable computational times for both the offline and online parts.

7.3.3. An Eco-routing algorithm for HEVs under traffic conditions

In [12], an extension of the bi-level optimization for the energy management of hybrid electric vehicles (HEVs) proposed above to the eco-routing problem is presented. Using the knowledge of traffic conditions over the entire road network, we search both the optimal path and state of charge trajectory. This problem results in finding the shortest path on a weighted graph whose nodes are (position, state of charge) pairs for the vehicle, the edge cost being evaluated thanks to the cost maps from optimization at the 'micro' level of a bi-level decomposition. The error due to the discretization of the state of charge is proven to be linear if the cost maps are Lipschitz. The classical $A*$ algorithm is used to solve the problem, with a heuristic based on a lower bound of the energy needed to complete the travel. The eco-routing method is validated by numerical simulations and compared to the fastest path on a synthetic road network.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. IPL

8.1.1.1. Algae in Silico

Inria Project Lab ALGAE IN SILICO (2014-2018) was dedicated to provide an integrated platform for numerical simulation of microalgae “from genes to industrial process“. Commands joined the project in 2017 to tackle the optimization aspects. Our previous collaborations with teams Modemic and Biocore on bioreactors [27], [15] have been renewed in this framework.

8.1.1.2. Cosy

Inria Project Lab COSY (started in 2017) aims at exploiting the potential of state-of-art biological modelling, control techniques, synthetic biology and experimental equipment to achieve a paradigm shift in control of microbial communities. More precisely, we plan to determine and implement control strategies to make heterogeneous communities diversify and interact in the most profitable manner. Study of yeast cells has started in collaboration with team Lifeware (G. Batt) in the framework of the PhD of V. Andreani, and is pursued in the Postdoc of D. Lunz (started Nov. 2019).

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

9.1.1.1. Member of the Conference Program Committees

- F. Bonnans: PGMO Days, EDF’Lab Palaiseau,Dec. 3-4, 2019.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- F. Bonnans: Associate Editor: Math. & Appl. / Annals of the Academy of Romanian Scientists (AOSR)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master :
F. Bonnans: *Numerical analysis of partial differential equations arising in finance and stochastic control*, 18h, M2, Ecole Polytechnique and U. Paris 6, France.
F. Bonnans: *Optimal control of ordinary differential equations*, 15h, M2, Optimization master (U. Paris-Saclay) and Ensta, France.
A. Kröner : Optimal control of partial differential equations, 20h, M2, Optimization master (U. Paris-Saclay), France.
L. Pfeiffer: *Optimal control of ordinary differential equations*, 18h, M2, Optimization master (U. Paris-Saclay) and Ensta, France.
L. Pfeiffer: *Optimisation continue et combinatoire*, 17h, Ensta, France.

9.2.2. Supervision

Finished PhD : A. Le Rhun, Optimal and robust control of hybrid vehicles. Started September 2016 (IFPEN fellowship), finished December 2019, F. Bonnans and P. Martinon.

PhD in progress : G. Bonnet, Efficient schemes for the Hamilton-Jacobi-Bellman equation. Started Oct. 2018. F. Bonnans and J.-M. Mirebeau, LMO, U. Orsay.

PhD in progress : P. Lavigne, Mathematical study of economic equilibria for renewable energy sources. Started Oct. 2018. F. Bonnans and L. Pfeiffer.

9.3. Popularization

- F. Bonnans: codirection of a joint Allistene-Ancre commission (contribution to the national strategy for research), Numerics and Energy committee (2017-2019).
- F. Bonnans: Dimitrie Pompeiu Prize Committee (Academy of Romanian Scientists).

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] M. S. Aronna, J. F. Bonnans, A. Kröner.*Optimal control of PDEs in a complex space setting; application to the Schrödinger equation*, in "SIAM Journal on Control and Optimization", 2019, vol. 57, n° 2, p. 1390-1412 [*DOI :* 10.1137/17M1117653], https://hal.archives-ouvertes.fr/hal-01311421

[2] J. F. Bonnans, J. Gianatti, F. Silva.*On the time discretization of stochastic optimal control problems: the dynamic programming approach*, in "ESAIM: Control, Optimisation and Calculus of Variations", October 2019 [*DOI :* 10.1051/cocv/2018045], https://hal.inria.fr/hal-01474285

[3] J. F. Bonnans, S. Hadikhanloo, L. Pfeiffer.*Schauder Estimates for a Class of Potential Mean Field Games of Controls*, in "Applied Mathematics and Optimization", July 2019, 34, https://hal.inria.fr/hal-02048437

[4] T. Breiten, K. Kunisch, L. Pfeiffer.*Feedback Stabilization of the Two-Dimensional Navier–Stokes Equations by Value Function Approximation*, in "Applied Mathematics and Optimization", December 2019, vol. 80, n° 3, p. 599-641 [*DOI :* 10.1007/s00245-019-09586-x], https://hal.archives-ouvertes.fr/hal-02372083

[5] T. Breiten, K. Kunisch, L. Pfeiffer.*Taylor expansions of the value function associated with a bilinear optimal control problem*, in "Annales de l'Institut Henri Poincaré (C) Non Linear Analysis", August 2019, vol. 36, n° 5, p. 1361-1399 [*DOI :* 10.1016/j.anihpc.2019.01.001], https://hal.archives-ouvertes.fr/hal-02372091

[6] K. Kunisch, L. Pfeiffer.*The effect of the terminal penalty in receding horizon control for a class of stabilization problems*, in "ESAIM: Control, Optimisation and Calculus of Variations", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02372077

[7] A. Le Rhun, J. F. Bonnans, G. De Nunzio, T. Leroy, P. Martinon.*A stochastic data-based traffic model applied to vehicles energy consumption estimation*, in "IEEE Transactions on Intelligent Transportation Systems", 2019, forthcoming [*DOI :* 10.1109/TITS.2019.2923292], https://hal.inria.fr/hal-01774621

[8] C. Rommel, J. F. Bonnans, P. Martinon, B. Gregorutti.*Gaussian Mixture Penalty for Trajectory Optimization Problems*, in "Journal of Guidance, Control, and Dynamics", August 2019, vol. 42, n° 8, p. 1857–1862 [*DOI :* 10.2514/1.G003996], https://hal.inria.fr/hal-01819749

Other Publications

[9] J. F. Bonnans, G. Bonnet, J.-M. Mirebeau.*Monotone and second order consistent schemes for the Pucci and Monge-Ampere equations*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02383521

[10] J. F. BONNANS, J. GIANATTI.*Optimal Control of an Age-Structured System with State Constraints*, June 2019, working paper or preprint, https://hal.inria.fr/hal-02164310

[11] A. LE RHUN, J. F. BONNANS, G. DE NUNZIO, T. LEROY, P. MARTINON.*A bi-level energy management strategy for HEVs under probabilistic traffic conditions*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02278359

[12] A. LE RHUN, J. F. BONNANS, G. DE NUNZIO, T. LEROY, P. MARTINON.*An Eco-routing algorithm for HEVs under traffic conditions*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02356277

References in notes

[13] M. BARDI, I. CAPUZZO-DOLCETTA.*Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, Systems and Control: Foundations and Applications, Birkhäuser, Boston, 1997

[14] G. BARLES.*Solutions de viscosité des équations de Hamilton-Jacobi*, Mathématiques et Applications, Springer, Paris, 1994, vol. 17

[15] T. BAYEN, F. MAIRET, P. MARTINON, M. SEBBAH.*Analysis of a periodic optimal control problem connected to microalgae anaerobic digestion*, in "Optimal Control Applications and Methods", 2014 [*DOI :* 10.1002/OCA.2127], https://hal.archives-ouvertes.fr/hal-00860570

[16] G. BLISS.*Lectures on the Calculus of Variations*, University of Chicago Press, Chicago, Illinois, 1946

[17] J. F. BONNANS, J. LAURENT-VARIN.*Computation of order conditions for symplectic partitioned Runge-Kutta schemes with application to optimal control*, in "Numerische Mathematik", 2006, vol. 103, n° 1, p. 1–10

[18] J. F. BONNANS, E. OTTENWAELTER, H. ZIDANI.*Numerical schemes for the two dimensional second-order HJB equation*, in "ESAIM: M2AN", 2004, vol. 38, p. 723-735

[19] J. F. BONNANS, H. ZIDANI.*Consistency of generalized finite difference schemes for the stochastic HJB equation*, in "SIAM J. Numerical Analysis", 2003, vol. 41, p. 1008-1021

[20] A. E. BRYSON, Y.-C. HO.*Applied optimal control*, Hemisphere Publishing, New-York, 1975

[21] M. CRANDALL, P. LIONS.*Viscosity solutions of Hamilton Jacobi equations*, in "Bull. American Mathematical Society", 1983, vol. 277, p. 1–42

[22] W. HAGER.*Runge-Kutta methods in optimal control and the transformed adjoint system*, in "Numerische Mathematik", 2000, vol. 87, n° 2, p. 247–282

[23] A. IOFFE, V. TIHOMIROV.*Theory of Extremal Problems*, North-Holland Publishing Company, Amsterdam, 1979, Russian Edition: Nauka, Moscow, 1974

[24] J.-L. LAGRANGE.*Mécanique analytique*, Paris, 1788, reprinted by J. Gabay, 1989

[25] E. LEE, L. MARKUS.*Foundations of optimal control theory*, John Wiley, New York, 1967

[26] G. LEITMANN.*An introduction to optimal control*, Mc Graw Hill, New York, 1966

[27] R. MUÑOZ-TAMAYO, P. MARTINON, G. BOUGARAN, F. MAIRET, O. BERNARD.*Getting the most out of it: Optimal experiments for parameter estimation of microalgae growth models*, in "Journal of Process Control", 2014, vol. 24, n^o 6, p. 991 - 1001 [*DOI :* 10.1016/J.JPROCONT.2014.04.021], http://www.sciencedirect.com/science/article/pii/S095915241400122X

[28] L. PONTRYAGIN, V. BOLTYANSKI, R. GAMKRELIDZE, E. MICHTCHENKO.*The Mathematical Theory of Optimal Processes*, Wiley Interscience, New York, 1962

[29] L. YOUNG.*Lectures on the calculus of variations and optimal control theory*, W. B. Saunders Co., Philadelphia, 1969

Project-Team DATASHAPE

Understanding the shape of data

RESEARCH CENTERS
Saclay - Île-de-France
Sophia Antipolis - Méditerranée

THEME
Algorithmics, Computer Algebra and Cryptology

Table of contents

Project-Team DATASHAPE

Creation of the Team: 2016 January 01, updated into Project-Team: 2016 January 01

Keywords:

Computer Science and Digital Science:

A3. - Data and knowledge
A3.4. - Machine learning and statistics
A7.1. - Algorithms
A8. - Mathematics of computing
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology
A9. - Artificial intelligence

Other Research Topics and Application Domains:

B1. - Life sciences
B2. - Health
B5. - Industry of the future
B9. - Society and Knowledge
B9.5. - Sciences

1. Team, Visitors, External Collaborators

Research Scientists

Frédéric Chazal [Team leader, Inria, Senior Researcher, HDR]
Jean-Daniel Boissonnat [Inria, Senior Researcher, HDR]
David Cohen-Steiner [Inria, Researcher]
Kunal Dutta [Inria, Starting Research Position, until Jan 2019]
Marc Glisse [Inria, Researcher]
Miroslav Kramar [Inria, Advanced Research Position, until Aug 2019]
Clément Maria [Inria, Researcher]
Steve Oudot [Inria, Researcher, HDR]
Mathijs Wintraecken [Inria, Starting Research Position, until Jan 2019]

Faculty Members

Clément Levrard [Univ Denis Diderot, Associate Professor, until Aug 2019]
Maksims Ovsjanikovs [École polytechnique, Professor]

External Collaborators

Clément Levrard [Univ Denis Diderot, from Sep 2019]
Bertrand Michel [Univ Pierre et Marie Curie]
Louis Pujol [Univ Paris-Sud, Oct 2019]
Georg Wilding [University of Groningen, from Sep 2019 until Nov 2019]
Claire Brecheteau [Ecole Centrale Nantes]

Technical Staff

Thomas Bonis [Inria, Engineer]
Vincent Rouvreau [Inria, Engineer]

PhD Students

Bertrand Beaufils [Sysnav, PhD Student, granted by CIFRE]

Nicolas Berkouk [Inria, PhD Student]
Denys Bulavka [Inria, PhD Student, until Apr 2019]
Alex Delalande [Inria, PhD Student, from Dec 2019]
Vincent Divol [Univ Paris-Sud, PhD Student]
Siargey Kachanovich [Inria, PhD Student, from Apr 2019 until May 2019]
Théo Lacombe [École polytechnique, PhD Student]
Etienne Lasalle [Univ Paris-Sud, PhD Student, from Sep 2019]
Siddharth Pritam [Inria, PhD Student]
Louis Pujol [Univ Paris-Sud, PhD Student, from Nov 2019]
Owen Rouille [Inria, PhD Student]
Raphael Tinarrage [École Normale Supérieure de Cachan, PhD Student]

Post-Doctoral Fellows

Hariprasad Kannan [Inria, Post-Doctoral Fellow]
Jisu Kim [Inria, Post-Doctoral Fellow]
Martin Royer [Inria, Post-Doctoral Fellow]

Visiting Scientists

Yoann Barszezak [École Normale Supérieure de Cachan, from Mar 2019 until Jul 2019]
Arijit Ghosh [Indian Statistical Institute, from Aug 2019 until Sep 2019]
Yuichi Ike [Fujitsu Laboratories, from May 2019 until Jul 2019]
Jacob Leygonie [Oxford, from Jul 2019 until Sep 2019]
Elchanan Solomon [Brown, until Jan 2019]

Administrative Assistants

Florence Barbara [Inria, Administrative Assistant, until May 2019]
Bahar Carabetta [Inria, Administrative Assistant]
Sophie Honnorat [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

DataShape is a research project in Topological Data Analysis (TDA), a recent field whose aim is to uncover, understand and exploit the topological and geometric structure underlying complex and possibly high dimensional data. The DATASHAPE project gathers a unique variety of expertise that allows it to embrace the mathematical, statistical, algorithmic and applied aspects of the field in a common framework ranging from fundamental theoretical studies to experimental research and software development.

The expected output of DATASHAPE is two-fold. First, we intend to set-up and develop the mathematical, statistical and algorithmic foundations of Topological and Geometric Data Analysis. Second, we intend to develop the Gudhi platform in order to provide an efficient state-of-the-art toolbox for the understanding of the topology and geometry of data.

3. Research Program

3.1. Algorithmic aspects of topological and geometric data analysis

TDA requires to construct and manipulate appropriate representations of complex and high dimensional shapes. A major difficulty comes from the fact that the complexity of data structures and algorithms used to approximate shapes rapidly grows as the dimensionality increases, which makes them intractable in high dimensions. We focus our research on simplicial complexes which offer a convenient representation of general shapes and generalize graphs and triangulations. Our work includes the study of simplicial complexes with good approximation properties and the design of compact data structures to represent them.

In low dimensions, effective shape reconstruction techniques exist that can provide precise geometric approximations very efficiently and under reasonable sampling conditions. Extending those techniques to higher dimensions as is required in the context of TDA is problematic since almost all methods in low dimensions rely on the computation of a subdivision of the ambient space. A direct extension of those methods would immediately lead to algorithms whose complexities depend exponentially on the ambient dimension, which is prohibitive in most applications. A first direction to by-pass the curse of dimensionality is to develop algorithms whose complexities depend on the intrinsic dimension of the data (which most of the time is small although unknown) rather than on the dimension of the ambient space. Another direction is to resort to cruder approximations that only captures the homotopy type or the homology of the sampled shape. The recent theory of persistent homology provides a powerful and robust tool to study the homology of sampled spaces in a stable way.

3.2. Statistical aspects of topological and geometric data analysis

The wide variety of larger and larger available data - often corrupted by noise and outliers - requires to consider the statistical properties of their topological and geometric features and to propose new relevant statistical models for their study.

There exist various statistical and machine learning methods intending to uncover the geometric structure of data. Beyond manifold learning and dimensionality reduction approaches that generally do not allow to assert the relevance of the inferred topological and geometric features and are not well-suited for the analysis of complex topological structures, set estimation methods intend to estimate, from random samples, a set around which the data is concentrated. In these methods, that include support and manifold estimation, principal curves/manifolds and their various generalizations to name a few, the estimation problems are usually considered under losses, such as Hausdorff distance or symmetric difference, that are not sensitive to the topology of the estimated sets, preventing these tools to directly infer topological or geometric information.

Regarding purely topological features, the statistical estimation of homology or homotopy type of compact subsets of Euclidean spaces, has only been considered recently, most of the time under the quite restrictive assumption that the data are randomly sampled from smooth manifolds.

In a more general setting, with the emergence of new geometric inference tools based on the study of distance functions and algebraic topology tools such as persistent homology, computational topology has recently seen an important development offering a new set of methods to infer relevant topological and geometric features of data sampled in general metric spaces. The use of these tools remains widely heuristic and until recently there were only a few preliminary results establishing connections between geometric inference, persistent homology and statistics. However, this direction has attracted a lot of attention over the last three years. In particular, stability properties and new representations of persistent homology information have led to very promising results to which the DATASHAPE members have significantly contributed. These preliminary results open many perspectives and research directions that need to be explored.

Our goal is to build on our first statistical results in TDA to develop the mathematical foundations of Statistical Topological and Geometric Data Analysis. Combined with the other objectives, our ultimate goal is to provide a well-founded and effective statistical toolbox for the understanding of topology and geometry of data.

3.3. Topological approach for multimodal data processing

Due to their geometric nature, multimodal data (images, video, 3D shapes, etc.) are of particular interest for the techniques we develop. Our goal is to establish a rigorous framework in which data having different representations can all be processed, mapped and exploited jointly. This requires adapting our tools and sometimes developing entirely new or specialized approaches.

The choice of multimedia data is motivated primarily by the fact that the amount of such data is steadily growing (with e.g. video streaming accounting for nearly two thirds of peak North-American Internet traffic, and almost half a billion images being posted on social networks each day), while at the same time it poses significant challenges in designing informative notions of (dis)-similarity as standard metrics (e.g. Euclidean distances between points) are not relevant.

3.4. Experimental research and software development

We develop a high quality open source software platform called GUDHI which is becoming a reference in geometric and topological data analysis in high dimensions. The goal is not to provide code tailored to the numerous potential applications but rather to provide the central data structures and algorithms that underlie applications in geometric and topological data analysis.

The development of the GUDHI platform also serves to benchmark and optimize new algorithmic solutions resulting from our theoretical work. Such development necessitates a whole line of research on software architecture and interface design, heuristics and fine-tuning optimization, robustness and arithmetic issues, and visualization. We aim at providing a full programming environment following the same recipes that made up the success story of the CGAL library, the reference library in computational geometry.

Some of the algorithms implemented on the platform will also be interfaced to other software platform, such as the R software [0] for statistical computing, and languages such as Python in order to make them usable in combination with other data analysis and machine learning tools. A first attempt in this direction has been done with the creation of an R package called TDA in collaboration with the group of Larry Wasserman at Carnegie Mellon University (Inria Associated team CATS) that already includes some functionalities of the GUDHI library and implements some joint results between our team and the CMU team. A similar interface with the Python language is also considered a priority. To go even further towards helping users, we will provide utilities that perform the most common tasks without requiring any programming at all.

4. New Software and Platforms

4.1. GUDHI

Geometric Understanding in Higher Dimensions

KEYWORDS: Computational geometry - Topology

SCIENTIFIC DESCRIPTION: The current release of the GUDHI library includes: – Data structures to represent, construct and manipulate simplicial and cubical complexes. – Algorithms to compute simplicial complexes from point cloud data. – Algorithms to compute persistent homology and multi-field persistent homology. – Simplification methods via implicit representations.

FUNCTIONAL DESCRIPTION: The GUDHI open source library will provide the central data structures and algorithms that underly applications in geometry understanding in higher dimensions. It is intended to both help the development of new algorithmic solutions inside and outside the project, and to facilitate the transfer of results in applied fields.

NEWS OF THE YEAR: - Cover complex - Representation of persistence diagrams - Cech complex - weighted periodic 3d alpha-complex - sparse Rips complex - debian / docker / conda-forge packages

- Participants: Clément Maria, François Godi, David Salinas, Jean-Daniel Boissonnat, Marc Glisse, Mariette Yvinec, Pawel Dlotko, Siargey Kachanovich, Vincent Rouvreau, Mathieu Carrière and Bertrand Michel
- Contact: Jean-Daniel Boissonnat
- URL: https://gudhi.inria.fr/

4.2. CGAL module: interval arithmetics

KEYWORD: Arithmetic

[0] https://www.r-project.org/

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) provides an efficient number type for intervals of double and the corresponding arithmetic operations. It is used in the evaluation of geometric predicates for a first quick computation, which either provides the result with guarantees, or rarely answers that more precision is needed.

RELEASE FUNCTIONAL DESCRIPTION: Partial rewrite to take advantage of SIMD instructions on recent x86 processors.

- Contact: Marc Glisse
- URL: https://www.cgal.org/

4.3. CGAL module: interface to Boost.Multiprecision

KEYWORD: Arithmetic

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) makes it possible to use some number types from Boost.Multiprecision in CGAL.

- Author: Marc Glisse
- Contact: Marc Glisse
- URL: https://www.cgal.org/

4.4. Module CGAL: New dD Geometry Kernel

KEYWORD: Computational geometry

FUNCTIONAL DESCRIPTION: This package of CGAL (Computational Geometry Algorithms Library http://www.cgal.org) provides the basic geometric types (point, vector, etc) and operations (orientation test, etc) used by geometric algorithms in arbitrary dimension. It uses filters for efficient exact predicates.

RELEASE FUNCTIONAL DESCRIPTION: New kernel with lazy exact constructions.

- Author: Marc Glisse
- Contact: Marc Glisse
- URL: http://www.cgal.org/

5. New Results

5.1. Algorithmic aspects of topological and geometric data analysis

5.1.1. Sampling and Meshing Submanifolds

Participants: Jean-Daniel Boissonnat, Siargey Kachanovich.

In collaboration with Mathijs Wintraecken (IST Autria).

This work [41], [11] presents a rather simple tracing algorithm to sample and mesh an m-dimensional submanifold of $\mathbb{R}^d$ for arbitrary m and d. We extend the work of Dobkin et al. to submanifolds of arbitrary dimension and codimension. The algorithm is practical and has been thoroughly investigated from both theoretical and experimental perspectives. The paper provides a full description and analysis of the data structure and of the tracing algorithm. The main contributions are : 1. We unify and complement the knowledge about Coxeter and Freudenthal-Kuhn triangulations. 2. We introduce an elegant and compact data structure to store Coxeter or Freudenthal-Kuhn triangulations and describe output sensitive algorithms to compute faces and cofaces or any simplex in the triangulation. 3. We present a manifold tracing algorithm based on the above data structure. We provide a detailled complexity analysis along with experimental results that show that the algorithm can handle cases that are far ahead of the state-of-the-art.

5.1.2. Topological correctness of PL-approximations of isomanifolds

Participant: Jean-Daniel Boissonnat.

In collaboration with Mathijs Wintraecken (IST Autria).

Isomanifolds are the generalization of isosurfaces to arbitrary dimension and codimension, i.e. manifolds defined as the zero set of some multivariate multivalued function $f: \mathbb{R}^d \to \mathbb{R}^{d-n}$. A natural (and efficient) way to approximate an isomanifold is to consider its Piecewise-Linear (PL) approximation based on a triangulation $\mathcal{T}$ of the ambient space $\mathbb{R}^d$. In this paper [43], we give conditions under which the PL-approximation of an isomanifold is topologically equivalent to the isomanifold. The conditions are easy to satisfy in the sense that they can always be met by taking a sufficiently fine triangulation $\mathcal{T}$. This contrasts with previous results on the triangulation of manifolds where, in arbitrary dimensions, delicate perturbations are needed to guarantee topological correctness, which leads to strong limitations in practice. We further give a bound on the Fréchet distance between the original isomanifold and its PL-approximation. Finally we show analogous results for the PL-approximation of an isomanifold with boundary.

5.1.3. Dimensionality Reduction for k-Distance Applied to Persistent Homology

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Shreya Arya (Duke University)

Given a set P of n points and a constant k, we are interested in computing the persistent homology of the Čech filtration of P for the k-distance, and investigate the effectiveness of dimensionality reduction for this problem, answering an open question of Sheehy [*Proc. SoCG, 2014*] [38]. We first show using the Johnson-Lindenstrauss lemma, that the persistent homology can be preserved up to a $(1 \pm \epsilon)$ factor while reducing dimensionality to $O(k \log n/\varepsilon^2)$. Our main result shows that the target dimension can be improved to $O(\log n/\varepsilon^2)$ under a reasonable and naturally occuring condition. The proof involves a multi-dimensional variant of the Hanson-Wright inequality for subgaussian quadratic forms and works when the random matrices are used for the Johnson-Lindenstrauss mapping are subgaussian. This includes the Gaussian matrices of Indyk-Motwani, the sparse random matrices of Achlioptas and the Ailon-Chazelle fast Johnson-Lindenstrauss transform. To provide evidence that our condition encompasses quite general situations, we show that it is satisfied when the points are independently distributed (i) in $\mathbb{R}^D$ under a subgaussian distribution, or (ii) on a spherical shell in $\mathbb{R}^D$ with a minimum angular separation, using Gershgorin's theorem. Our results also show that the JL-mapping preserves up to a $(1 \pm \epsilon)$ factor, the Rips and Delaunay filtrations for the k-distance, as well as the Čech filtration for the approximate k-distance of Buchet et al.

5.1.4. Edge Collapse and Persistence of Flag Complexes

Participants: Jean-Daniel Boissonnat, Siddharth Pritam.

In this article [42], we extend the notions of dominated vertex and strong collapse of a simplicial complex as introduced by J. Barmak and E. Miniam adn build on the initial success of [30]. We say that a simplex (of any dimension) is dominated if its link is a simplicial cone. Domination of edges appear to be very powerful and we study it in the case of flag complexes in more detail. We show that edge collapse (removal of dominated edges) in a flag complex can be performed using only the 1-skeleton of the complex. Furthermore, the residual complex is a flag complex as well. Next we show that, similar to the case of strong collapses, we can use edge collapses to reduce a flag filtration $\mathcal{F}$ to a smaller flag filtration $\mathcal{F}^c$ with the same persistence. Here again, we only use the 1-skeletons of the complexes. The resulting method to compute $\mathcal{F}^c$ is simple and extremely efficient and, when used as a preprocessing for Persistence Computation, leads to gains of several orders of magnitude wrt the state-of-the-art methods (including our previous approach using strong collapse). The method is exact, irrespective of dimension, and improves performance of Persistence Computation even in low dimensions. This is demonstrated by numerous experiments on publicly available data.

5.1.5. DTM-based Filtrations

Participants: Frédéric Chazal, Marc Glisse, Raphael Tinarrage.

In collaboration with Anai, Hirokazu and Ike, Yuichi and Inakoshi, Hiroya and Umeda, Yuhei (Fujitsu Labs).

Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris-Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In [15], we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions, and extends some previous work on the approximation of such functions.

5.1.6. Recovering the homology of immersed manifolds

Participant: Raphael Tinarrage.

Given a sample of an abstract manifold immersed in some Euclidean space, in [57], we describe a way to recover the singular homology of the original manifold. It consists in estimating its tangent bundle -seen as subset of another Euclidean space- in a measure theoretic point of view, and in applying measure-based filtrations for persistent homology. The construction we propose is consistent and stable, and does not involve the knowledge of the dimension of the manifold.

5.1.7. Regular triangulations as lexicographic optimal chains

Participant: David Cohen-Steiner.

In collaboration with André Lieutier and Julien Vuillamy (Dassault Systèmes).

We introduce [46] a total order on n-simplices in the n-Euclidean space for which the support of the lexicographic-minimal chain with the convex hull boundary as boundary constraint is precisely the n-dimensional Delaunay triangulation, or in a more general setting, the regular triangulation of a set of weighted points. This new characterization of regular and Delaunay triangulations is motivated by its possible generalization to submanifold triangulations as well as the recent development of polynomial-time triangulation algorithms taking advantage of this order.

5.1.8. Discrete Morse Theory for Computing Zigzag Persistence

Participant: Clément Maria.

In collaboration with Hannah Schreiber (Graz University of Technology, Austria)

We introduce a framework to simplify zigzag filtrations of general complexes using discrete Morse theory, in order to accelerate the computation of zigzag persistence. Zigzag persistence is a powerful algebraic generalization of persistent homology. However, its computation is much slower in practice, and the usual optimization techniques cannot be used to compute it. Our approach is different in that it preprocesses the filtration before computation. Using discrete Morse theory, we get a much smaller zigzag filtration with same persistence. The new filtration contains general complexes. We introduce new update procedures to modify on the fly the algebraic data (the zigzag persistence matrix) under the new combinatorial changes induced by the Morse reduction. Our approach is significantly faster in practice [35].

5.1.9. Computing Persistent Homology with Various Coefficient Fields in a Single Pass

Participants: Jean-Daniel Boissonnat, Clément Maria.

This article [18] introduces an algorithm to compute the persistent homology of a filtered complex with various coefficient fields in a single matrix reduction. The algorithm is output-sensitive in the total number of distinct persistent homological features in the diagrams for the different coefficient fields. This computation allows us to infer the prime divisors of the torsion coefficients of the integral homology groups of the topological space at any scale, hence furnishing a more informative description of topology than persistence in a single coefficient field. We provide theoretical complexity analysis as well as detailed experimental results. The code is part of the Gudhi software library.

5.1.10. Exact computation of the matching distance on 2-parameter persistence modules

Participant: Steve Oudot.

In collaboration with Michael Kerber (T.U. Graz) and Michael Lesnick (SUNY).

The matching distance is a pseudometric on multi-parameter persistence modules, defined in terms of the weighted bottleneck distance on the restriction of the modules to affine lines. It is known that this distance is stable in a reasonable sense, and can be efficiently approximated, which makes it a promising tool for practical applications. In [31] we show that in the 2-parameter setting, the matching distance can be computed exactly in polynomial time. Our approach subdivides the space of affine lines into regions, via a line arrangement. In each region, the matching distance restricts to a simple analytic function, whose maximum is easily computed. As a byproduct, our analysis establishes that the matching distance is a rational number, if the bigrades of the input modules are rational.

5.1.11. Decomposition of exact pfd persistence bimodules

Participant: Steve Oudot.

In collaboration with Jérémy Cochoy (Symphonia).

In [24] we identify a certain class of persistence modules indexed over $\mathbb{R}^2$ that are decomposable into direct sums of indecomposable summands called blocks. The conditions on the modules are that they are both pointwise finite-dimensional (pfd) and exact. Our proof follows the same scheme as the one for pfd persistence modules indexed over $\mathbb{R}$, yet it departs from it at key stages due to the product order not being a total order on $\mathbb{R}^2$, which leaves some important gaps open. These gaps are filled in using more direct arguments. Our work is motivated primarily by the study of interlevel-sets persistence, although the proposed results reach beyond that setting.

5.1.12. Level-sets persistence and sheaf theory

Participants: Nicolas Berkouk, Steve Oudot.

In collaboration with Grégory Ginot (Paris 13).

In [39] we provide an explicit connection between level-sets persistence and derived sheaf theory over the real line. In particular we construct a functor from 2-parameter persistence modules to sheaves over R, as well as a functor in the other direction. We also observe that the 2-parameter persistence modules arising from the level sets of Morse functions carry extra structure that we call a Mayer-Vietoris system. We prove classification, barcode decomposition, and stability theorems for these Mayer-Vietoris systems, and we show that the aforementioned functors establish a pseudo-isometric equivalence of categories between derived constructible sheaves with the convolution or (derived) bottleneck distance and the interleaving distance of strictly pointwise finite-dimensional Mayer-Vietoris systems. Ultimately, our results provide a functorial equivalence between level-sets persistence and derived pushforward for continuous real-valued functions.

5.1.13. Intrinsic Interleaving Distance for Merge Trees

Participant: Steve Oudot.

In collaboration with Ellen Gasparovic (Union College), Elizabeth Munch (Michigan State), Katharine Turner (Australian National University), Bei Wang (Utah), and Yusu Wang (Ohio-State).

Merge trees are a type of graph-based topological summary that tracks the evolution of connected components in the sublevel sets of scalar functions. They enjoy widespread applications in data analysis and scientific visualization. In [49] we consider the problem of comparing two merge trees via the notion of interleaving distance in the metric space setting. We investigate various theoretical properties of such a metric. In particular, we show that the interleaving distance is intrinsic on the space of labeled merge trees and provide an algorithm to construct metric 1-centers for collections of labeled merge trees. We further prove that the intrinsic property of the interleaving distance also holds for the space of unlabeled merge trees. Our results are a first step toward performing statistics on graph-based topological summaries.

5.2. Statistical aspects of topological and geometric data analysis

5.2.1. Estimating the Reach of a Manifold

Participants: Frédéric Chazal, Jisu Kim, Bertrand Michel.

In collaboration with E. Aamari (Univ. Paris-Diderot), A. Rinaldo, L. Wasserman (Carnegie Mellon University).

In [13], various problems in manifold estimation make use of a quantity called the reach, denoted by τ_M, which is a measure of the regularity of the manifold. This paper is the first investigation into the problem of how to estimate the reach. First, we study the geometry of the reach through an approximation perspective. We derive new geometric results on the reach for submanifolds without boundary. An estimator $\widehat{\tau}$ of τ_M is proposed in an oracle framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud X_n, $\widehat{\tau}(X_n)$ is showed to achieve uniform expected loss bounds over a C^3-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

5.2.2. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature

Participant: Claire Brecheteau.

In [20], we introduce the notion of DTM-signature, a measure on $\mathbb{R}$ that can be associated to any metric-measure space. This signature is based on the function distance to a measure (DTM) introduced in 2009 by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, that is bounded above by the Gromov-Wasserstein distance. This pseudo-metric is used to build a statistical test of isomorphism between two metric-measure spaces, from the observation of two N-samples.

The test is based on subsampling methods and comes with theoretical guarantees. It is proven to be of the correct level asymptotically. Also, when the measures are supported on compact subsets of $\mathbb{R}^d$, rates of convergence are derived for the $L1$-Wasserstein distance between the distribution of the test statistic and its subsampling approximation. These rates depend on some parameter $\rho > 1$. In addition, we prove that the power is bounded above by $\exp(-CN1/\rho)$, with C proportional to the square of the aforementioned pseudo-metric between the metric-measure spaces. Under some geometrical assumptions, we also derive lower bounds for this pseudo-metric.

An algorithm is proposed for the implementation of this statistical test, and its performance is compared to the performance of other methods through numerical experiments.

5.2.3. On the choice of weight functions for linear representations of persistence diagrams

Participant: Vincent Divol.

In collaboration with Wolfgang Polonik (UC Davis).

Persistence diagrams are efficient descriptors of the topology of a point cloud. As they do not naturally belong to a Hilbert space, standard statistical methods cannot be directly applied to them. Instead, feature maps (or representations) are commonly used for the analysis. A large class of feature maps, which we call linear, depends on some weight functions, the choice of which is a critical issue. An important criterion to choose a weight function is to ensure stability of the feature maps with respect to Wasserstein distances on diagrams. In [21], we improve known results on the stability of such maps, and extend it to general weight functions. We also address the choice of the weight function by considering an asymptotic setting; assume that $\mathbb{X}_n$ is an i.i.d. sample from a density on $[0,1]^d$. For the Č ech and Rips filtrations, we characterize the weight functions for which the corresponding feature maps converge as n approaches infinity, and by doing so, we prove laws of large numbers for the total persistences of such diagrams. Those two approaches (stability and convergence) lead to the same simple heuristic for tuning weight functions: if the data lies near a d-dimensional manifold, then a sensible choice of weight function is the persistence to the power α with $\alpha \geq d$.

5.2.4. Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport

Participants: Vincent Divol, Théo Lacombe.

Despite the obvious similarities between the metrics used in topological data analysis and those of optimal transport, an optimal-transport based formalism to study persistence diagrams and similar topological descriptors has yet to come. In [48], by considering the space of persistence diagrams as a measure space, and by observing that its metrics can be expressed as solutions of optimal partial transport problems, we introduce a generalization of persistence diagrams, namely Radon measures supported on the upper half plane. Such measures naturally appear in topological data analysis when considering continuous representations of persistence diagrams (e.g. persistence surfaces) but also as limits for laws of large numbers on persistence diagrams or as expectations of probability distributions on the persistence diagrams space. We study the topological properties of this new space, which will also hold for the closed subspace of persistence diagrams. New results include a characterization of convergence with respect to transport metrics, the existence of Fréchet means for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the usefulness of this framework to study random persistence diagrams by providing several statistical results made meaningful thanks to this new formalism.

5.3. Topological approach for multimodal data processing

5.3.1. A General Neural Network Architecture for Persistence Diagrams and Graph Classification

Participants: Frédéric Chazal, Théo Lacombe, Martin Royer.

In collaboration with Mathieu Carrière (Colombia Univ.) and Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

Persistence diagrams, the most common descriptors of Topological Data Analysis, encode topological properties of data and have already proved pivotal in many different applications of data science. However, since the (metric) space of persistence diagrams is not Hilbert, they end up being difficult inputs for most Machine Learning techniques. To address this concern, several vectorization methods have been put forward that embed persistence diagrams into either finite-dimensional Euclidean space or (implicit) infinite dimensional Hilbert space with kernels. In [44], we focus on persistence diagrams built on top of graphs. Relying on extended persistence theory and the so-called heat kernel signature, we show how graphs can be encoded by (extended) persistence diagrams in a provably stable way. We then propose a general and versatile framework for learning vectorizations of persistence diagrams, which encompasses most of the vectorization techniques used in the literature. We finally showcase the experimental strength of our setup by achieving competitive scores on classification tasks on real-life graph datasets.

5.3.2. Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks

Participant: Frédéric Chazal.

In collaboration with Umeda Yuhei and Meryll Dindin (Fujitsu Labs).

In [47], we present an innovative and generic deep learning approach to monitor heart conditions from ECG signals.We focus our attention on both the detection and classification of abnormal heartbeats, known as arrhythmia. We strongly insist on generalization throughout the construction of a deep-learning model that turns out to be effective for new unseen patient. The novelty of our approach relies on the use of topological data analysis as basis of our multichannel architecture, to diminish the bias due to individual differences. We show that our structure reaches the performances of the state-of-the-art methods regarding arrhythmia detection and classification.

5.3.3. ATOL: Automatic Topologically-Oriented Learning

Participants: Frédéric Chazal, Martin Royer.

In collaboration with Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

There are abundant cases for using Topological Data Analysis (TDA) in a learning context, but robust topological information commonly comes in the form of a set of persistence diagrams, objects that by nature are uneasy to affix to a generic machine learning framework. In [56], we introduce a vectorisation method for diagrams that allows to collect information from topological descriptors into a format fit for machine learning tools. Based on a few observations, the method is learned and tailored to discriminate the various important plane regions a diagram is set into. With this tool one can automatically augment any sort of machine learning problem with access to a TDA method, enhance performances, construct features reflecting underlying changes in topological behaviour. The proposed methodology comes with only high level tuning parameters such as the encoding budget for topological features. We provide an open-access, ready-to-use implementation and notebook. We showcase the strengths and versatility of our approach on a number of applications. From emulous and modern graph collections to a highly topological synthetic dynamical orbits data, we prove that the method matches or beats the state-of-the-art in encoding persistence diagrams to solve hard problems. We then apply our method in the context of an industrial, difficult time-series regression problem and show the approach to be relevant.

5.3.4. Inverse Problems in Topological Persistence: a Survey

Participant: Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

In [27] we review the literature on inverse problems in topological persistence theory.The first half of the survey is concerned with the question of surjectivity, i.e. the existence of rightinverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlightthe tools and theorems that underlie these advances, and direct the reader's attention to openproblems, both theoretical and applied.

5.3.5. Intrinsic Topological Transforms via the Distance Kernel Embedding

Participants: Clément Maria, Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

Topological transforms are parametrized families of topological invariants, which, by analogy with transforms in signal processing, are much more discriminative than single measurements. The first two topological transforms to be defined were the Persistent Homology Transform and Euler Characteristic Transform, both of which apply to shapes embedded in Euclidean space. The contribution of this work [54] is to define topological transforms that depend only on the intrinsic geometry of a shape, and hence are invariant to the choice of embedding. To that end, given an abstract metric measure space, we define an integral operator whose eigenfunctions are used to compute sublevel set persistent homology. We demonstrate that this operator, which we call the distance kernel operator, enjoys desirable stability properties, and that its spectrum and eigenfunctions concisely encode the large-scale geometry of our metric measure space. We then define a number of topological transforms using the eigenfunctions of this operator, and observe that these transforms inherit many of the stability and injectivity properties of the distance kernel operator.

5.3.6. A Framework for Differential Calculus on Persistence Barcodes

Participant: Steve Oudot.

In collaboration with Jacob Leygonie and Ulrike Tillmann (Oxford).

In [52], we define notions of differentiability for maps from and to the space of persistence barcodes. Inspired by the theory of diffeological spaces, the proposed framework uses lifts to the space of ordered barcodes, from which derivatives can be computed. The two derived notions of differentiability (respectively from and to the space of barcodes) combine together naturally to produce a chain rule that enables the use of gradient descent for objective functions factoring through the space of barcodes. We illustrate the versatility of this framework by showing how it can be used to analyze the smoothness of various parametrized families of filtrations arising in topological data analysis.

5.4. Experimental research and software development

5.4.1. Robust Stride Detector from Ankle-Mounted Inertial Sensors for Pedestrian Navigation and Activity Recognition with Machine Learning Approaches

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [16], a stride detector algorithm combined with a technique inspired by zero velocity update (ZUPT) is proposed to reconstruct the trajectory of a pedestrian from an ankle-mounted inertial device. This innovative approach is based on sensor alignment and machine learning. It is able to detect 100% of both normal walking strides and more than 97% of atypical strides such as small steps, side steps, and backward walking that existing methods can hardly detect. This approach is also more robust in critical situations, when for example the wearer is sitting and moving the ankle or when the wearer is bicycling (less than two false detected strides per hour on average). As a consequence, the algorithm proposed for trajectory reconstruction achieves much better performances than existing methods for daily life contexts, in particular in narrow areas such as in a house. The computed stride trajectory contains essential information for recognizing the activity (atypical stride, walking, running, and stairs). For this task, we adopt a machine learning approach based on descriptors of these trajectories, which is shown to be robust to a large of variety of gaits. We tested our algorithm on recordings of healthy adults and children, achieving more than 99% success. The algorithm also achieved more than 97by children suffering from movement disorders. Compared to most algorithms in the literature, this original method does not use a fixed-size sliding window but infers this last in an adaptive way

5.4.2. *Robust pedestrian trajectory reconstruction from inertial sensor*

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [28], a strides detection algorithm combined with a technique inspired by Zero Velocity Update (ZUPT) is proposed using inertial sensors worn on the ankle. This innovative approach based on a sensors alignment and machine learning can detect both normal walking strides and atypical strides such as small steps, side steps and backward walking that existing methods struggle to detect. As a consequence, the trajectory reconstruction achieves better performances in daily life contexts for example, where a lot of these kinds of strides are performed in narrow areas such as in a house. It is also robust in critical situations, when for example the wearer is sitting and moving the ankle or bicycling, while most algorithms in the literature would wrongly detect strides and produce error in the trajectory reconstruction by generating movements.Our algorithm is evaluated on more than 7800 strides from seven different subjects performing several activities. We validated the trajectory reconstruction during motion capture sessions by analyzing the stride length. Finally, we tested the algorithm in a challenging situation by plotting the computed trajectory on the building map of an 5 hours and 30 minutes office worker recording.

5.5. Algorithmic and Combinatorial Aspects of Low Dimensional Topology

5.5.1. *Treewidth, crushing and hyperbolic volume*

Participant: Clément Maria.

In collaboration with Jessica S. Purcell (Monash University, Australia)

The treewidth of a 3-manifold triangulation plays an important role in algorithmic 3-manifold theory, and so it is useful to find bounds on the tree-width in terms of other properties of the manifold. In [26], we prove that there exists a universal constant c such that any closed hyperbolic 3-manifold admits a triangulation of tree-width at most the product of c and the volume. The converse is not true: we show there exists a sequence of hyperbolic 3-manifolds of bounded tree-width but volume approaching infinity. Along the way, we prove that crushing a normal surface in a triangulation does not increase the carving-width, and hence crushing any number of normal surfaces in a triangulation affects tree-width by at most a constant multiple.

5.5.2. *Parameterized complexity of quantum knot invariants*

Participant: Clément Maria.

In [53], we give a general fixed parameter tractable algorithm to compute quantum invariants of links presented by diagrams, whose complexity is singly exponential in the carving-width (or the tree-width) of the diagram. In particular, we get a $O(N^{3/2\mathrm{cw}}\mathrm{poly}(n))$ time algorithm to compute any Reshetikhin-Turaev invariant-derived from a simple Lie algebra g of a link presented by a planar diagram with n crossings and carving-width cw, and whose components are coloured with g-modules of dimension at most N. For example, this includes the Nth-coloured Jones polynomial and the Nth-coloured HOMFLYPT polynomial.

5.6. Miscellaneous

5.6.1. Material Coherence from Trajectories via Burau Eigenanalysis of Braids

Participant: David Cohen-Steiner.

In collaboration with Melissa Yeung and Mathieu Desbrun (Caltech).

In this paper [58], we provide a numerical tool to study material coherence from a set of 2D Lagrangian trajectories sampling a dynamical system, i.e., from the motion of passive tracers. We show that eigenvectors of the Burau representation of a topological braid derived from the trajectories have levelsets corresponding to components of the Nielsen-Thurston decomposition of the dynamical system. One can thus detect and identify clusters of space-time trajectories corresponding to coherent regions of the dynamical system by solving an eigenvalue problem. Unlike previous methods, the scalable computational complexity of our braid-based approach allows the analysis of large amounts of trajectories. Studying two-dimensional flows and their induced transport and mixing properties is key to geophysical studies of atmospheric and oceanic processes. However, one often has only sparse tracer trajectories (e.g., positions of buoys in time) to infer the overall flow geometry. Fortunately, topological methods based on the theory of braid groups have recently been proposed to extract structures from such a sparse set of trajectories by measuring their entan-glement. This braid viewpoint offers sound foundations for the definition of coherent structures. Yet, there has been only limited efforts in developing practical tools that can leverage topological properties for the efficient analysis of flow structures: handling a larger number of tra-jectories remains computationally challenging. We contribute a new and simple computational tool to extract Lagrangian structures from sparse trajectories by noting that the eigenstructure of the Burau matrix representation of a braid of particle trajectories can be used to reveal coherent regions of the flows. Detection of clusters of space-time trajectories corresponding to coherent regions of the dynamical system can thus be achieved by solving a simple eigenvalue problem. This paper establishes the theoretical foundations behind this braid eigenanalysis approach, along with numerical validations on various flows.

5.6.2. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space

Participants: Alex Delalande, Frédéric Chazal.

In collaboration with Quentin Mérigot (Institut de Mathématiques d'Orsay).

In [55], we study an explicit embedding of the set of probability measures into a Hilbert space, defined using optimal transport maps from a reference probability density. This embedding linearizes to some extent the 2-Wasserstein space, and enables the direct use of generic supervised and unsupervised learning algorithms on measure data. Our main result is that the embedding is (bi-)Holder continuous, when the reference density is uniform over a convex set, and can be equivalently phrased as a dimension-independent Hölder-stability results for optimal transport maps.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

- Collaboration with Sysnav, a French SME with world leading expertise in navigation and geopositioning in extreme environments, on TDA, geometric approaches and machine learning for the analysis of movements of pedestrians and patients equipped with inetial sensors (CIFRE PhD of Bertrand Beaufils).
- Research collaboration with Fujitsu on the development of new TDA methods and tools for Machine learning and Artificial Intelligence (started in Dec 2017).
- Research collaboration with MetaFora on the development of new TDA-based and statistical methods for the analysis of cytometric data (started in Nov. 2019).

6.2. Bilateral Grants with Industry

- DATASHAPE and Sysnav have been selected for the ANR/DGA Challenge MALIN (funding: 700 kEuros) on pedestrian motion reconstruction in severe environments (without GPS access).

7. Partnerships and Cooperations

7.1. Regional Initiatives

Mini course on "Sheaf Theory and Topological Data Analysis" taught by Rodrigo Cordoniu (Nice University) at Inria Sophia Antipolis — 8 weeks, 2h per week, Feb 2019 to Apr 2019.

7.2. National Initiatives

7.2.1. ANR

7.2.1.1. ANR ASPAG

Participant: Marc Glisse.

- Acronym : ASPAG.

- Type : ANR blanc.

- Title : Analysis and Probabilistic Simulations of Geometric Algorithms.

- Coordinator : Olivier Devillers (équipe Inria Gamble).

- Duration : 4 years from January 2018 to December 2021.

- Others Partners: Inria Gamble, LPSM, LABRI, Université de Rouen, IECL, Université du Littoral Côte d'Opale, Telecom ParisTech, Université Paris X (Modal'X), LAMA, Université de Poitiers, Université de Bourgogne.

- Abstract:

The analysis and processing of geometric data has become routine in a variety of human activities ranging from computer-aided design in manufacturing to the tracking of animal trajectories in ecology or geographic information systems in GPS navigation devices. Geometric algorithms and probabilistic geometric models are crucial to the treatment of all this geometric data, yet the current available knowledge is in various ways much too limited: many models are far from matching real data, and the analyses are not always relevant in practical contexts. One of the reasons for this state of affairs is that the breadth of expertise required is spread among different scientific communities (computational geometry, analysis of algorithms and stochastic geometry) that historically had very little interaction. The Aspag project brings together experts of these communities to address the problem of geometric data. We will more specifically work on the following three interdependent directions.

(1) Dependent point sets: One of the main issues of most models is the core assumption that the data points are independent and follow the same underlying distribution. Although this may be relevant in some contexts, the independence assumption is too strong for many applications.

(2) Simulation of geometric structures: The phenomena studied in (1) involve intricate random geometric structures subject to new models or constraints. A natural first step would be to build up our understanding and identify plausible conjectures through simulation. Perhaps surprisingly, the tools for an effective simulation of such complex geometric systems still need to be developed.

(3) Understanding geometric algorithms: the analysis of algorithm is an essential step in assessing the strengths and weaknesses of algorithmic principles, and is crucial to guide the choices made when designing a complex data processing pipeline. Any analysis must strike a balance between realism and tractability; the current analyses of many geometric algorithms are notoriously unrealistic. Aside from the purely scientific objectives, one of the main goals of Aspag is to bring the communities closer in the long term. As a consequence, the funding of the project is crucial to ensure that the members of the consortium will be able to interact on a very regular basis, a necessary condition for significant progress on the above challenges.

- See also: https://members.loria.fr/Olivier.Devillers/aspag/

7.3. International Research Visitors

7.3.1. Visits of International Scientists

- Arijit Ghosh, Indian Statistical Institute, Kolkata, India (September 2019)
- Ramsay Dyer Berkeley Publishing (September 2019)
- Mathijs Wintraecken, IST Austria (September and October 2019)

7.3.1.1. Internships

- Alex Delalande, Centrale-Supelec, (May-October 2019).

7.3.1.2. Research Stays Abroad

- Martin Royer, Fujitsu Laboratories, Tokyo, 2 months.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

- F. Chazal was co-organiser of the Geometric Data Analysis conference at the Stevanovich Center for Financial Mathematics, Chicago, May 2019.
- F. Chazal was co-organiser of the conference New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

8.1.2. Scientific Events: Selection

8.1.2.1. Chair of Conference Program Committees

- Steve Oudot was chair of the program committee for the Young Researchers Forum at the International Symposium on Computational Geometry (SoCG) 2019.

8.1.2.2. Member of the Conference Program Committees

- Jean-Daniel Boissonnat was a member of the Organizing Committee of the International Conference on Curves and Surfaces 2019.
- Clément Maria was a member of the Program Committee of the International Symposium on Computational Geometry (SoCG), Portland, USA, June 2019.
- David Cohen-Steiner was a member of the Program Committees of SGP'19 and SMI'19.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

Jean-Daniel Boissonnat is a member of the Editorial Board of *Journal of the ACM*, *Discrete and Computational Geometry*, *International Journal on Computational Geometry and Applications*.

Frédéric Chazal is a member of the Editorial Board of *SIAM Journal on Imaging Sciences*, *Discrete and Computational Geometry (Springer)*, *Graphical Models (Elsevier), and Journal of Applied and Computational Topology (Springer)*.

Frédéric Chazal is a member of the Scientific Board of Journal of Applied and Computational Topology (Springer).

Steve Oudot is a member of the Editorial Board of *Journal of Computational Geometry*.

8.1.4. Invited Talks

Frédéric Chazal, Geometry of Big Data, Institute of Pure and Applied Mathematics at UCLA, Los Angeles, April 2019.

Frédéric Chazal, Workshop on Data Science, Fundaçao Getulio Vargas, Rio de Janeiro, April 2019.

Frédéric Chazal, SUMTOPO 2019, Topology in Data Science, Johannesburg, June 2019.

Frédéric Chazal, New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

Frédéric Chazal, Seminar@System X, Saclay, June 2019.

Clément Maria, Monash University, Australia.

Clément Maria, New Horizons in Computational Geometry and Topology, Sophia-Antipolis, September 2019.

Clément Maria, Applied Algebraic Topology Research Network Seminar, October 2019.

Clément Maria, Dagstuhl Seminar 19352 "Computation in Low-Dimensional Geometry and Topology", August 2019.

Clément Maria, 2ème Demi-journée Pole Calcul - Topologie et Géométrie du Calcul - LIS Marseille, June 2019.

Steve Oudot, Workshop on computational applications of quiver representations: TDA and QPA, Bielefeld University, April 2019.

Steve Oudot, Minisymposium on Algebraic Geometry in Topological Data Analysis, SIAM Conference on Applied Algebraic Geometry, Bern, July 2019.

Steve Oudot, Summer Conference on Topology and its Applications, Johannesburg, July 2019.

Steve Oudot, Spires workshop: from theory to applications of TDA, Oxford, September 2019.

8.1.5. Leadership within the Scientific Community

Frédéric Chazal is co-responsible, with S. Arlot (Paris-Sud Univ.), of the "programme Maths-STIC" of the Labex Fondation Mathématique Jacques Hadamard (FMJH).

Frédéric Chazal is a member of the "Comité de pilotage" of the SIGMA group at SMAI.

Steve Oudot is co-head with Luca Castelli Aleardi (École polytechnique) of the GT GeoAlgo within the GdR-IM.

8.1.6. Research Administration

Marc Glisse is president of the CDT at Inria Saclay.

Steve Oudot is vice-president of the Commission Scientifique at Inria Saclay.

Clément Maria is a member of the CDT at Inria Sophia Antipolis-Méditerranée.

Clément Maria, responsable RAweb pour Datashape.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master: Frédéric Chazal and Quentin Mérigot, Analyse Topologique des Données, 30h eq-TD, Université Paris-Sud, France.

Master: Marc Glisse and Clément Maria, Computational Geometry Learning, 36h eq-TD, M2, MPRI, France.

Master: Frédéric Cazals and Frédéric Chazal, Geometric Methods for Data Analysis, 30h eq-TD, M1, École Centrale Paris, France.

Master: Frédéric Chazal and Julien Tierny, Topological Data Analysis, 38h eq-TD, M2, Mathématiques, Vision, Apprentissage (MVA), ENS Paris-Saclay, France.

Course on TDA at the Statistics with Geometry and Topology Conference 2019, 4hCM, Toulouse, France.

Master: Steve Oudot, Topological data analysis, 45h eq-TD, M1, École polytechnique, France.

Master: Steve Oudot, Data Analysis: geometry and topology in arbitrary dimensions, 24h eq-TD, M2, graduate program in Artificial Intelligence & Advanced Visual Computing, École polytechnique, France.

Undergrad-Master: Steve Oudot, Algorithms for data analysis in C++, 22.5h eq-TD, L3/M1, École polytechnique, France.

8.2.2. Supervision

PhD : Alba Chiara de Vitis, Kernel Methods for High Dimensional Data Analysis [12]. Université Côte d'Azur. May 28, 2019. Jean-Daniel Boissonnat and David Cohen-Steiner.

PhD : Sergei Kachanovich, Meshing submanifolds using Coxeter triangulations [11]. Université Côte d'Azur. October 23, 2019. Jean-Daniel Boissonnat.

PhD: Yitchzak Solomon, Inverse problems in topological data analysis, Brown University. Defended in May 2019. Jeff Brock (Yale University) and Steve Oudot.

PhD in progress: Raphaël Tinarrage, Persistence and stability of nerves in measured metric spaces for Topological Data Analysis, started September 1st, 2017, Frédéric Chazal and Marc Glisse.

PhD in progress: Louis Pujol, Partitionnement de données cytométriques, started Novermber 1st, 2019, Pascal Massart and Marc Glisse.

PhD in progress: Bertrand Beaufils, Méthodes topologiques et apprentissage statistique pour l'actimétrie du piéton à partir de données de mouvement, Frédéric Chazal and Bertrand Michel (Ecole Centrale de Nantes).

PhD in progress: Vincent Divol, statistical aspects of TDA, started September 1st, 2017, Frédéric Chazal and Pascal Massart (LMO).

PhD in progress: Etienne Lasalle, TDA for graph data, started September 1st, 2019, Frédéric Chazal and Pascal Massart (LMO).

PhD in progress: Alex Delalande, Measure embedding with Optimal Transport and applications in Machine Learning, started December 1st, 2019, Frédéric Chazal and Quentin Mérigot (LMO).

PhD in progress: Owen Rouillé, started September 2018, co-advised by C. Maria and J-D. Boissonnat.

PhD in progress: Nicolas Berkouk, Categorification of topological graph structures, started November 1st, 2016, Steve Oudot.

PhD in progress: Théo Lacombe, Statistics for persistence diagrams using optimal transport, started October 1st, 2017. Marco Cuturi (ENSAE & Google Brain) and Steve Oudot.

8.2.3. Juries

Clément Maria was a member of the jury attributing the Gilles Kahn PhD award, from the SIF and the Academy of Science, Nov. 2019.

Steve Oudot was reviewer and committee member for the PhD defence of Hannah Schreiber, October 2019.

8.3. Popularization

8.3.1. Interventions

- Frédéric Chazal, Un exemple de dispositif dédié aux études cliniques. Seminar on AI, Haute Autorité de Santé, September 2019.

9. Bibliography

Major publications by the team in recent years

[1] D. Attali, U. Bauer, O. Devillers, M. Glisse, A. Lieutier.*Homological Reconstruction and Simplification in R3*, in "Computational Geometry", 2014 [*DOI :* 10.1016/j.comgeo.2014.08.010], https://hal.archives-ouvertes.fr/hal-01132440

[2] J.-D. Boissonnat, R. Dyer, A. Ghosh.*Delaunay Triangulation of Manifolds*, in "Foundations of Computational Mathematics", 2017, vol. 45, 38 [*DOI :* 10.1007/s10208-017-9344-1], https://hal.inria.fr/hal-01509888

[3] J.-D. Boissonnat, R. Dyer, A. Ghosh, S. Y. Oudot.*Only distances are required to reconstruct submanifolds*, in "Computational Geometry", 2017, vol. 66, p. 32 - 67 [*DOI :* 10.1016/j.comgeo.2017.08.001], https://hal.inria.fr/hal-01583086

[4] J.-D. Boissonnat, K. C. Srikanta, S. Tavenas.*Building Efficient and Compact Data Structures for Simplicial Complexe*, in "Algorithmica", September 2016 [*DOI :* 10.1007/s00453-016-0207-y], https://hal.inria.fr/hal-01364648

[5] F. Chazal, D. Cohen-Steiner, A. Lieutier.*A Sampling Theory for Compact Sets in Euclidean Space*, in "Discrete Comput. Geom.", 2009, vol. 41, n^{o} 3, p. 461–479, http://dx.doi.org/10.1007/s00454-009-9144-8

[6] F. Chazal, D. Cohen-Steiner, Q. Mérigot.*Geometric Inference for Measures based on Distance Functions*, in "Foundations of Computational Mathematics", 2011, vol. 11, n^{o} 6, p. 733-751, RR-6930 [*DOI :* 10.1007/s10208-011-9098-0], http://hal.inria.fr/inria-00383685

[7] F. Chazal, S. Y. Oudot, M. Glisse, V. De Silva.*The Structure and Stability of Persistence Modules*, SpringerBriefs in Mathematics, Springer Verlag, 2016, VII, 116, https://hal.inria.fr/hal-01330678

[8] L. J. Guibas, S. Y. Oudot, P. Skraba, F. Chazal.*Persistence-Based Clustering in Riemannian Manifolds*, in "Journal of the ACM", November 2013, vol. 60, n^{o} 6, 38, http://hal.archives-ouvertes.fr/hal-00923563

[9] M. Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez, M. Desbrun.*Variance-Minimizing Transport Plans for Inter-surface Mapping*, in "ACM Transactions on Graphics", 2017, vol. 36, 14 [*DOI :* 10.1145/3072959.3073671], https://hal.inria.fr/hal-01519006

[10] S. Y. Oudot.*Persistence Theory: From Quiver Representations to Data Analysis*, Mathematical Surveys and Monographs, American Mathematical Society, 2015, n^{o} 209, 218, https://hal.inria.fr/hal-01247501

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. KACHANOVICH.*Meshing submanifolds using Coxeter triangulations*, Université Côte d'Azur, October 2019, https://hal.inria.fr/tel-02419148

[12] A. C. DE VITIS.*Kernel Methods for High Dimensional Data Analysis*, Université Côte d'Azur, May 2019, https://hal.inria.fr/tel-02419727

Articles in International Peer-Reviewed Journal

[13] E. AAMARI, J. KIM, F. CHAZAL, B. MICHEL, A. RINALDO, L. WASSERMAN.*Estimating the Reach of a Manifold*, in "Electronic journal of statistics ", April 2019, https://arxiv.org/abs/1705.04565 [*DOI :* 10.1214/19-EJS1551], https://hal.archives-ouvertes.fr/hal-01521955

[14] E. AAMARI, C. LEVRARD.*Non-Asymptotic Rates for Manifold, Tangent Space, and Curvature Estimation*, in "Annals of Statistics", 2019, vol. 47, n^o 1, https://arxiv.org/abs/1705.00989 [*DOI :* 10.1214/18-AOS1685], https://hal.archives-ouvertes.fr/hal-01516032

[15] H. ANAI, F. CHAZAL, M. GLISSE, Y. IKE, H. INAKOSHI, R. TINARRAGE, Y. UMEDA.*DTM-based Filtrations*, in "Abel Symposia", 2019, https://arxiv.org/abs/1811.04757 , forthcoming, https://hal.archives-ouvertes.fr/hal-01919562

[16] B. BEAUFILS, F. CHAZAL, M. GRELET, B. MICHEL.*Robust Stride Detector from Ankle-Mounted Inertial Sensors for Pedestrian Navigation and Activity Recognition with Machine Learning Approaches*, in "Sensors", October 2019, vol. 19, n^o 20, 4491 [*DOI :* 10.3390/s19204491], https://hal.inria.fr/hal-02365094

[17] J.-D. BOISSONNAT, A. LIEUTIER, M. WINTRAECKEN.*The reach, metric distortion, geodesic convexity and the variation of tangent spaces*, in "Journal of Applied and Computational Topology", June 2019 [*DOI :* 10.1007/s41468-019-00029-8], https://hal.inria.fr/hal-02266258

[18] J.-D. BOISSONNAT, C. MARIA.*Computing Persistent Homology with Various Coefficient Fields in a Single Pass*, in "Journal of Applied and Computational Topology", September 2019, vol. 3, n^o 1-2, 16, https://arxiv.org/abs/2001.02960 [*DOI :* 10.1007/s41468-019-00025-y], https://hal.inria.fr/hal-00922572

[19] J.-D. BOISSONNAT, M. ROUXEL-LABBÉ, M. WINTRAECKEN.*Anisotropic triangulations via discrete Riemannian Voronoi diagrams*, in "SIAM Journal on Computing", 2019, https://hal.inria.fr/hal-02419460

[20] C. BRÉCHETEAU.*A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature*, in "Electronic journal of statistics ", 2019, MSC: Primary 62G10; secondary 62G09, forthcoming [*DOI :* 10.1214/154957804100000000], https://hal.inria.fr/hal-01426331

[21] V. DIVOL, W. POLONIK.*On the choice of weight functions for linear representations of persistence diagrams*, in "Journal of Applied and Computational Topology", August 2019, https://arxiv.org/abs/1807.03678 [*DOI :* 10.1007/s41468-019-00032-z], https://hal.inria.fr/hal-01833660

[22] K. DUTTA, A. GHOSH, B. JARTOUX, N. MUSTAFA.*Shallow packings, semialgebraic set systems, Macbeath regions, and polynomial partitioning*, in "Discrete and Computational Geometry", 2019, vol. 61, n^o 4, p. 756-777 [*DOI :* 10.1007/S00454-019-00075-0], https://hal.archives-ouvertes.fr/hal-02316975

[23] S. HARKER, M. KRAMÁR, R. LEVANGER, K. MISCHAIKOW.*A Comparison Framework for Interleaved Persistence Modules*, in "Journal of Applied and Computational Topology", 2019 [*DOI :* 10.1007/S41468-019-00026-X], https://hal.inria.fr/hal-01757092

[24] C. JÉRÉMY, S. Y. OUDOT.*Decomposition of exact pfd persistence bimodules*, in "Discrete and Computational Geometry", December 2019, https://arxiv.org/abs/1605.09726 , https://hal.inria.fr/hal-01359312

[25] J. KIM, A. RINALDO, L. WASSERMAN.*Minimax Rates for Estimating the Dimension of a Manifold*, in "Journal of Computational Geometry", 2019, vol. 10, n^o 1, https://arxiv.org/abs/1605.01011 - 54 pages, 11 figures [*DOI :* 10.20382/JOCG.V10I1A3], https://hal.inria.fr/hal-02425684

[26] C. MARIA, J. PURCELL.*Treewidth, crushing and hyperbolic volume*, in "Algebraic and Geometric Topology", 2019, vol. 19, n^o 5, p. 2625-2652, https://arxiv.org/abs/1805.02357 [*DOI :* 10.2140/AGT.2019.19.2625], https://hal.archives-ouvertes.fr/hal-02429712

[27] S. Y. OUDOT, E. SOLOMON.*Inverse Problems in Topological Persistence: a Survey*, in "Abel Symposia", 2019, https://arxiv.org/abs/1810.10813 , forthcoming, https://hal.inria.fr/hal-01966676

International Conferences with Proceedings

[28] B. BEAUFILS, F. CHAZAL, M. GRELET, B. MICHEL.*Robust pedestrian trajectory reconstruction from inertial sensor*, in "IPIN 2019 - 10th International Conference on Indoor Positioning and Indoor Navigation", Pisa, Italy, 2019, https://hal.inria.fr/hal-02271580

[29] J.-D. BOISSONNAT, O. DEVILLERS, K. DUTTA, M. GLISSE.*Randomized incremental construction of Delaunay triangulations of nice point sets*, in "ESA 2019 - 27th Annual European Symposium on Algorithms", Munich, Germany, September 2019 [*DOI :* 10.4230/LIPICS.ESA.2019.22], https://hal.inria.fr/hal-02185566

[30] J.-D. BOISSONNAT, S. PRITAM.*Computing Persistent Homology of Flag Complexes via Strong Collapses*, in "SoCG 2019 - International Symposium on Computational geometry", Portland, United States, April 2019 [*DOI :* 10.4230/LIPICS.SOCG.2019.55], https://hal.inria.fr/hal-02078311

[31] M. KERBER, M. LESNICK, S. Y. OUDOT.*Exact computation of the matching distance on 2-parameter persistence modules*, in "International Symposium on Computational Geometry", Portland, Oregon, United States, 2019, https://arxiv.org/abs/1812.09085 , https://hal.inria.fr/hal-01966666

[32] J. KIM, J. SHIN, A. RINALDO, L. WASSERMAN.*Uniform Convergence of the Kernel Density Estimator Adaptive to Intrinsic Volume Dimension*, in "Thirty-sixth International Conference on Machine Learning (ICML 2019)", Long Beach, United States, Proceedings of Thirty-sixth International Conference on Machine Learning (ICML 2019), 2019, vol. 97, https://arxiv.org/abs/1810.05935 - 51 pages, https://hal.inria.fr/hal-02404295

Conferences without Proceedings

[33] H. ANAI, F. CHAZAL, M. GLISSE, Y. IKE, H. INAKOSHI, R. TINARRAGE, Y. UMEDA.*DTM-based Filtrations*, in "SoCG 2019 - 35th International Symposium on Computational Geometry", Portland, United

States, June 2019, https://arxiv.org/abs/1811.04757 - This version is an extended abstract. The complete version of the paper, including proofs and additional comments, can be found at https://arxiv.org/abs/1811.04757 [*DOI :* 10.4230/LIPIcs.SoCG.2019.58], https://hal.archives-ouvertes.fr/hal-02093445

Scientific Books (or Scientific Book chapters)

[34] H. KANNAN, N. KOMODAKIS, N. PARAGIOS.*Tighter continuous relaxations for MAP inference in discrete MRFs: A survey*, in "Handbook of Numerical Analysis: Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2, Volume 20, pages 351 - 400, Elsevier, https://www.sciencedirect.com/science/article/pii/S1570865919300092", Elsevier, October 2019 [*DOI :* 10.1016/BS.HNA.2019.06.001], https://hal.inria.fr/hal-02432873

[35] C. MARIA, H. SCHREIBER.*Discrete Morse Theory for Computing Zigzag Persistence*, in "Algorithms and Data Structures - 16th International Symposium, WADS 2019, Edmonton, AB, Canada, August 5-7, 2019, Proceedings", Z. FRIGGSTAD, J.-R. SACK, M. R. SALAVATIPOUR (editors), Springer, July 2019, p. 538-552, https://arxiv.org/abs/1807.05172 [*DOI :* 10.1007/978-3-030-24766-9_39], https://hal.inria.fr/hal-01971682

Research Reports

[36] J.-D. BOISSONNAT, O. DEVILLERS, K. DUTTA, M. GLISSE.*Randomized incremental construction of Delaunay triangulations of nice point sets*, Inria, 2019, https://hal.inria.fr/hal-01950119

[37] O. DEVILLERS, P. DUCHON, M. GLISSE, X. GOAOC.*On Order Types of Random Point Sets*, Inria, 2019, https://arxiv.org/abs/1812.08525 , https://hal.inria.fr/hal-01962093

Other Publications

[38] S. ARYA, J.-D. BOISSONNAT, K. DUTTA.*Dimensionality Reduction for k-Distance Applied to Persistent Homology*, December 2019, working paper or preprint, https://hal.inria.fr/hal-01950051

[39] N. BERKOUK, G. GINOT, S. Y. OUDOT.*Level-sets persistence and sheaf theory*, December 2019, https://arxiv.org/abs/1907.09759 - working paper or preprint, https://hal.inria.fr/hal-02425597

[40] J.-D. BOISSONNAT, R. DYER, A. GHOSH, A. LIEUTIER, M. WINTRAECKEN.*Local conditions for triangulating submanifolds of Euclidean space*, July 2019, working paper or preprint, https://hal.inria.fr/hal-02267620

[41] J.-D. BOISSONNAT, S. KACHANOVICH, M. WINTRAECKEN.*Sampling and Meshing Submanifolds in High Dimension*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02386169

[42] J.-D. BOISSONNAT, S. PRITAM.*Edge Collapse and Persistence of Flag Complexes*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02395227

[43] J.-D. BOISSONNAT, M. WINTRAECKEN.*The topological correctness of PL-approximations of isomanifolds*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02386193

[44] M. CARRIERE, F. CHAZAL, Y. IKE, T. LACOMBE, M. ROYER, Y. UMEDA.*A General Neural Network Architecture for Persistence Diagrams and Graph Classification*, April 2019, working paper or preprint, https://hal.inria.fr/hal-02105788

[45] D. COHEN-STEINER, A. C. DE VITIS.*Spectral Properties of Radial Kernels and Clustering in High Dimensions*, January 2020, working paper or preprint, https://hal.inria.fr/hal-01969956

[46] D. COHEN-STEINER, A. LIEUTIER, J. VUILLAMY.*Regular triangulations as lexicographic optimal chains*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02391285

[47] M. DINDIN, Y. UMEDA, F. CHAZAL.*Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks*, June 2019, https://arxiv.org/abs/1906.05795 - 7 pages, 4 figures, https://hal.inria.fr/hal-02155849

[48] V. DIVOL, T. LACOMBE.*Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport*, April 2019, https://arxiv.org/abs/1901.03048 - working paper or preprint, https://hal.inria.fr/hal-01970466

[49] E. GASPAROVIC, E. MUNCH, S. Y. OUDOT, K. TURNER, B. WANG, Y. WANG.*Intrinsic Interleaving Distance for Merge Trees*, December 2019, https://arxiv.org/abs/1908.00063 - working paper or preprint, https://hal.inria.fr/hal-02425600

[50] J. KIM, J. SHIN, F. CHAZAL, A. RINALDO, L. WASSERMAN.*Homotopy Reconstruction via the Cech Complex and the Rips Complex*, December 2019, https://arxiv.org/abs/1903.06955 - 41 pages, 4 figures, https://hal.inria.fr/hal-02425686

[51] M. KRAMÁR, L. KOVALCINOVA, K. MISCHAIKOW, L. KONDIC.*Quantitative Measure of Memory Loss in Complex Spatio-Temporal Systems*, March 2019, working paper or preprint, https://hal.inria.fr/hal-02065027

[52] J. LEYGONIE, S. Y. OUDOT, U. TILLMANN.*A Framework for Differential Calculus on Persistence Barcodes*, October 2019, https://arxiv.org/abs/1910.00960 - 30 pages, 5 figures [*DOI :* 10.00960], https://hal.archives-ouvertes.fr/hal-02304300

[53] C. MARIA.*Parameterized complexity of quantum knot invariants*, January 2020, https://arxiv.org/abs/1910.00477 - working paper or preprint [*DOI :* 10.00477], https://hal.archives-ouvertes.fr/hal-02429767

[54] C. MARIA, S. Y. OUDOT, E. SOLOMON.*Intrinsic Topological Transforms via the Distance Kernel Embedding*, December 2019, https://arxiv.org/abs/1912.02225 - 50 pages, https://hal.inria.fr/hal-02425598

[55] Q. MÉRIGOT, A. DELALANDE, F. CHAZAL.*Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space*, October 2019, https://arxiv.org/abs/1910.05954 - 21 pages [*DOI :* 10.05954], https://hal.archives-ouvertes.fr/hal-02315998

[56] M. ROYER, F. CHAZAL, Y. IKE, Y. UMEDA.*ATOL: Automatic Topologically-Oriented Learning*, August 2019, https://arxiv.org/abs/1909.13472 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02296513

[57] R. TINARRAGE.*Recovering the homology of immersed manifolds*, December 2019, https://arxiv.org/abs/1912.03033 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02396261

[58] M. YEUNG, D. COHEN-STEINER, M. DESBRUN.*Material Coherence from Trajectories via Burau Eigenanalysis of Braids*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02295987

Project-Team DEDUCTEAM

DEDUCTEAM

IN COLLABORATION WITH: Laboratoire specification et vérification (LSV)

IN PARTNERSHIP WITH:
ENS Paris-Saclay

RESEARCH CENTER
Saclay - Île-de-France

THEME
Proofs and Verification

Table of contents

Project-Team DEDUCTEAM

Creation of the Team: 2011 December 01, updated into Project-Team: 2017 January 01

Keywords:

Computer Science and Digital Science:
A2.1.4. - Functional programming
A2.1.11. - Proof languages
A2.4.3. - Proofs
A3.1.1. - Modeling, representation
A7. - Theory of computation
A7.2. - Logic in Computer Science

Other Research Topics and Application Domains:
B7. - Transport and logistics

1. Team, Visitors, External Collaborators

Research Scientists
Gilles Dowek [Team leader, Inria, Senior Researcher, HDR]
Bruno Barras [Inria, Researcher]
Frédéric Blanqui [Inria, Researcher, HDR]
Valentin Blot [Inria, Researcher, from Mar 2019]
Jean-Pierre Jouannaud [University Paris-Saclay, Emeritus, HDR]

Faculty Members
Pablo Arrighi [Inria, Professor, from Sep 2019]
Guillaume Burel [ENSIIE, Associate Professor, until Aug 2019]

External Collaborators
Guillaume Burel [ENSIIE, from Sep 2019]
Catherine Dubois [ENSIIE]
Olivier Hermant [École Nationale Supérieure des Mines de Paris]

PhD Students
Mohamed Yacine El Haddad [CNRS, PhD Student]
Gaspard Ferey [École Nationale Supérieure des Mines de Paris , PhD Student]
Guillaume Genestier [École Normale Supérieure de Cachan, PhD Student]
Emilie Grienenberger [École Normale Supérieure de Cachan, PhD Student, from Oct 2019]
Gabriel Hondet [Inria, PhD Student, from Oct 2019]
François Thiré [École Normale Supérieure de Cachan, PhD Student, until Nov 2019]

Post-Doctoral Fellows
Michael Farber [Inria, Post-Doctoral Fellow, from Sep 2019]
Rehan Malak [CNRS, Post-Doctoral Fellow, from Nov 2019]
Étienne Miquey [CNRS, Post-Doctoral Fellow, from Nov 2019]
Franck Slama [Inria, Post-Doctoral Fellow, until Aug 2019]

Administrative Assistants
Adeline Lochet [Inria, Administrative Assistant, until May 2019]
Emmanuelle Perrot [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Objectives

The project-team investigates the design of logical frameworks, in order to ensure interoperability between proof systems, and to the development of system-independent proof libraries. To achieve these goals, we develop

- a logical framework DEDUKTI, where several theories can be expressed,
- tools to import proofs developed in external proof systems to DEDUKTI theories,
- tools to translate proofs from one DEDUKTI theory to another,
- tools to export proofs expressed in DEDUKTI theories to an external proof system,
- tools to prove the confluence, the termination, and the consistency of theories expressed in DEDUKTI,
- tools to develop proofs directly in DEDUKTI,
- an encyclopedia LOGIPEDIA of proofs expressed in various DEDUKTI theories.

2.2. History

The idea that systems such as Euclidean geometry or set theory should be expressed, not as independent systems, but in a logical framework appeared with the design of the first logical framework: predicate logic, in 1928. Later, several more powerful logical frameworks have been designed: λ-prolog, Isabelle, the Edinburgh logical framework, Pure type systems, and Deduction modulo theory.

The logical framework that we use is a simple λ-calculus with dependent types and rewrite rules, called the $\lambda\Pi$-calculus modulo theory, and also the Martin-Löf logical framework, and it generalizes all the mentioned frameworks. It is implemented in the system DEDUKTI.

The first version of DEDUKTI was developed in 2011 by Mathieu Boespflug [24]. From 2012 to 2015, new versions of DEDUKTI were developed and several theories were expressed in DEDUKTI, allowing to import proofs developed in MATITA (with the tool KRAJONO), HOL LIGHT (with the tool HOLIDE), FOCALIZE (with the tool FOCALIDE), IPROVER, and ZENON, totalizing several hundred of megabytes of proofs.

From 2015 to 2018, we focused on the translation of proofs from one DEDUKTI theory to another and to the exporting of proofs to other proof systems. In particular the MATITA arithmetic library has been translated to a much weaker theory: constructive simple type theory, allowing to export it to COQ, LEAN, PVS, HOL LIGHT, and ISABELLE/HOL. This led us to develop, in 2018, an online proof encyclopedia LOGIPEDIA, allowing to share and browse this library. We also focused on the development of new theories in DEDUKTI, and on an interactive theorem prover on top of DEDUKTI.

3. Research Program

3.1. Logical Frameworks

A thesis, which is at the root of our research effort, is that logical systems should be expressed as theories in a logical framework. As a consequence, proof-checking systems should not be focused on one theory, such as Simple type theory, Martin-Löf's type theory, or the Calculus of constructions, but should be theory independent. On the more theoretical side, the proof search algorithms, or the algorithmic interpretation of proofs should not depend on the theory in which proofs are expressed, but this theory should just be a parameter. This is for instance expressed in the title of our invited talk at ICALP 2012: *A theory independent Curry-De Bruijn-Howard correspondence* [25].

Various limits of Predicate logic have led to the development of various families of logical frameworks: λ-prolog and Isabelle have allowed terms containing free variables, the Edinburgh logical framework has allowed proofs to be expressed as λ-terms, Pure type systems have allowed propositions to be considered as terms, and Deduction modulo theory has allowed theories to be defined not only with axioms, but also with computation rules.

The $\lambda\Pi$-calculus modulo theory, that is implemented in the system DEDUKTI and that is a synthesis of the Edinburgh logical framework and of Deduction modulo theory, subsumes them all. Part of our research effort is focused on improving the $\lambda\Pi$-calculus modulo theory, for instance allowing to define congruences with associative and commutative rewriting. Another part of our research effort is focused on the automatic analysis of theories to prove their confluence, termination, and consistency either by pencil and paper proofs or automatically [4].

3.2. Interoperability and proof encyclopediae

Using a single prover to check proofs coming from different systems naturally leads to investigate how these proofs can be translated from one theory to another and used in a system different from the system in which they have been developed. This issue is of prime importance because developments in proof systems are getting bigger and, unlike other communities in computer science, the proof checking community has given little effort in the direction of standardization and interoperability.

For each proof, independently of the system in which it has been developed, we should be able to identify the systems in which it can be expressed. For instance, we have shown that many proofs developed in the MATITA prover did not use the full strength of the logic of MATITA and could be exported, for instance, to the systems of the HOL family, that are based on a weaker logic.

Rather than importing proofs from one system, transforming them, and exporting them to another system, we can use the same tools to develop system-independent proof encyclopedia called Logipedia. In such a library, each proof is labeled with the theories in which it can be expressed and so with the systems in which it can be used.

3.3. Interactive theorem proving

If our main goal with DEDUKTI is to import, transform, and export proofs developed in other systems, we also want to investigate how DEDUKTI can be used as the basis of an interactive theorem prover. This leads to two new scientific questions: first, how much can a tactic system be theory independent, and then how does rewriting extends the possibility to write tactics.

This has led to the development of a new version of DEDUKTI, which supports metavariables. Several tactics have been developed for this system, which are intended to help a human user to write proofs in our system instead of writing proof terms by hand. This work is a continuation of the previous work the team did on DEMON, which was an extension of DEDUKTI, whereas the support for interactive theorem proving is now native in DEDUKTI.

4. Application Domains

4.1. Interoperability

Our main impact applications, for instance to proofs of programs, or to air traffic control, are through our cooperation with other teams.

As a matter of fact, we view our work on interoperability and on the design of a formal proof encyclopedia as a service to the formal proof community.

5. New Software and Platforms

5.1. Autotheo

KEYWORD: Automated deduction

SCIENTIFIC DESCRIPTION: Transformation of axiomatic theories into rewriting systems that can be used by iProverModulo.

FUNCTIONAL DESCRIPTION: Autotheo is a tool that transforms axiomatic theories into polarized rewriting systems, thus making them usable in iProverModulo. It supports several strategies to orient the axioms, some of them being proved to be complete, in the sense that ordered polarized resolution modulo the resulting systems is refutationally complete, some others being merely heuristics. In practice, Autotheo takes a TPTP input file and produces an input file for iProverModulo.

NEWS OF THE YEAR: Maintenance.

- Participant: Guillaume Burel
- Partner: ENSIIE
- Contact: Guillaume Burel
- Publication: Consistency Implies Cut Admissibility
- URL: http://www.ensiie.fr/~guillaume.burel/blackandwhite_autotheo.html.en

5.2. CoLoR

Coq Library on Rewriting and termination

KEYWORDS: Coq - Formalisation

FUNCTIONAL DESCRIPTION: CoLoR is a Coq library on rewriting theory and termination. It provides many definitions and theorems on various mathematical structures (quasi-ordered sets, relations, ordered semi-rings, etc.), data structures (lists, vectors, matrices, polynomials, finite graphs), term structures (strings, first-order terms, lambda-terms, etc.), transformation techniques (dependency pairs, semantic labeling, etc.) and (non-)termination criteria (polynomial and matrix interpretations, recursive path ordering, computability closure, etc.).

- Authors: Frédéric Blanqui and Sébastien Hinderer
- Contact: Frédéric Blanqui
- Publications: CoLoR: a Coq library on well-founded rewrite relations and its application to the automated verification of termination certificates - Automated Verification of Termination Certificates - CoLoR: a Coq library on rewriting and termination
- URL: http://color.inria.fr/

5.3. Coqine

Coq In dEdukti

KEYWORDS: Higher-order logic - Formal methods - Proof

FUNCTIONAL DESCRIPTION: CoqInE is a plugin for the Coq software translating Coq proofs into Dedukti terms. It provides a Dedukti signature file faithfully encoding the underlying theory of Coq (or a sufficiently large subset of it). Current development is mostly focused on implementing support for Coq universe polymorphism. The generated ouput is meant to be type-checkable using the latest version of Dedukti.

- Contact: Guillaume Burel
- URL: http://www.ensiie.fr/~guillaume.burel/blackandwhite_coqInE.html.en

5.4. Dedukti

KEYWORD: Logical Framework

FUNCTIONAL DESCRIPTION: Dedukti is a proof-checker for the LambdaPi-calculus modulo. As it can be parametrized by an arbitrary set of rewrite rules, defining an equivalence relation, this calculus can express many different theories. Dedukti has been created for this purpose: to allow the interoperability of different theories.

Dedukti's core is based on the standard algorithm for type-checking semi-full pure type systems and implements a state-of-the-art reduction machine inspired from Matita's and modified to deal with rewrite rules.

Dedukti's input language features term declarations and definitions (opaque or not) and rewrite rule definitions. A basic module system allows the user to organize his project in different files and compile them separately.

Dedukti features matching modulo beta for a large class of patterns called Miller's patterns, allowing for more rewriting rules to be implemented in Dedukti.

NEWS OF THE YEAR: There has been a new release 2.6 in 2018. This release provides a better control on module loading, and a better log of rewrite steps.

- Participants: François Thiré, Gaspard Ferey, Guillaume Genestier and Rodolphe Lepigre
- Contact: François Thiré
- Publications: Dedukti:un vérificateur de preuves universel - Rewriting Modulo β in the $\lambda\Pi$-Calculus Modulo - Expressing theories in the $\lambda\Pi$-calculus modulo theory and in the Dedukti system
- URL: https://deducteam.github.io/

5.5. Holide

KEYWORD: Proof

FUNCTIONAL DESCRIPTION: Holide translates HOL proofs to Dedukti[OT] proofs, using the OpenTheory standard (common to HOL Light and HOL4). Dedukti[OT] being the encoding of OpenTheory in Dedukti.

- Contact: Guillaume Burel
- URL: http://deducteam.gforge.inria.fr/holide/

5.6. HOT

Higher-Order Termination

FUNCTIONAL DESCRIPTION: HOT is an automated termination prover for higher-order rewriting, based on the notion of computability closure.

- Contact: Frédéric Blanqui
- URL: http://rewriting.gforge.inria.fr/hot.html

5.7. iProver Modulo

KEYWORDS: Automated deduction - Automated theorem proving

SCIENTIFIC DESCRIPTION: Integration of ordered polarized resolution modulo theory into the prover iProver.

FUNCTIONAL DESCRIPTION: iProver Modulo is an extension of the automated theorem prover iProver originally developed by Konstantin Korovin at the University of Manchester. It implements ordered polarized resolution modulo theory, a refinement of the resolution method based on deduction modulo theory. It takes as input a proposition in predicate logic and a clausal rewriting system defining the theory in which the formula has to be proved. Normalization with respect to the term rewriting rules is performed very efficiently through translation into OCaml code, compilation and dynamic linking. Experiments have shown that ordered polarized resolution modulo dramatically improves proof search compared to using raw axioms.

NEWS OF THE YEAR: Maintenance of Dedukti output

- Participant: Guillaume Burel
- Partner: ENSIIE
- Contact: Guillaume Burel
- Publications: A Shallow Embedding of Resolution and Superposition Proofs into the ??-Calculus Modulo - Experimenting with deduction modulo
- URL: https://github.com/gburel/iProverModulo

5.8. mSAT

KEYWORD: Propositional logic

FUNCTIONAL DESCRIPTION: mSAT is a modular, proof-producing, SAT and SMT core based on Alt-Ergo Zero, written in OCaml. The solver accepts user-defined terms, formulas and theory, making it a good tool for experimenting. This tool produces resolution proofs as trees in which the leaves are user-defined proof of lemmas.

- Contact: Guillaume Bury
- Publication: mSAT:An OCaml SAT Solver
- URL: https://github.com/Gbury/mSAT

5.9. Rainbow

Termination certificate verifier

KEYWORDS: Demonstration - Code generation - Verification

FUNCTIONAL DESCRIPTION: Rainbow is a set of tools for automatically verifying the correctness of termination certificates expressed in the CPF format used in the annual international competition of termination tools. It contains: a tool xsd2coq for generating Coq data types for representing XML files valid with respect to some XML Schema, a tool xsd2ml for generating OCaml data types and functions for parsing XML files valid with respect to some XML Schema, a tool for translating a CPF file into a Coq script, and a standalone Coq certified tool for verifying the correctness of a CPF file.

- Author: Frédéric Blanqui
- Contact: Frédéric Blanqui
- Publications: Automated verification of termination certificates - Automated verification of termination certificates
- URL: http://color.inria.fr/rainbow.html

5.10. Krajono

KEYWORD: Proof

FUNCTIONAL DESCRIPTION: Krajono translates Matita proofs into Dedukti[CiC] (encoding of CiC in Dedukti) terms.

- Contact: François Thiré

5.11. archsat

KEYWORDS: Automated theorem proving - First-order logic - Propositional logic

FUNCTIONAL DESCRIPTION: Archsat is an automated theorem prover aimed at studying the integration of first-order theorem prover technologies, such as rewriting, into SMT solvers.

- Contact: Guillaume Bury
- URL: https://gforge.inria.fr/projects/archsat

5.12. lrat2dk

KEYWORDS: Automated theorem proving - Proof

FUNCTIONAL DESCRIPTION: Take as input a SAT proof trace in LRAT format, which can be obtained from the de facto standard format DRAT using drat-trim. Output a proof checkable by Dedukti, in a shallow encoding of propositional logic.

- Participant: Guillaume Burel
- Partner: ENSIIE
- Contact: Guillaume Burel
- URL: https://github.com/gburel/lrat2dk

5.13. ekstrakto

KEYWORDS: TPTP - TSTP - Proof assistant - Dedukti

FUNCTIONAL DESCRIPTION: Extracting TPTP problems from a TSTP trace. Proof reconstruction in Dedukti from TSTP trace.

- Contact: Mohamed Yacine El Haddad
- URL: https://github.com/elhaddadyacine/ekstrakto

5.14. SizeChangeTool

KEYWORDS: Rewriting systems - Proof assistant - Termination

FUNCTIONAL DESCRIPTION: A termination-checker for higher-order rewriting with dependent types. Took part in the Termination Competition 2018 (http://termination-portal.org/wiki/Termination_Competition_2018) in the "Higher-Order Rewriting (union Beta)" category.

- Partner: Mines ParisTech
- Contact: Guillaume Genestier
- URL: https://github.com/Deducteam/SizeChangeTool

5.15. Logipedia

KEYWORDS: Formal methods - Web Services - Logical Framework

FUNCTIONAL DESCRIPTION: Logipedia is composed of two distinct parts: 1) A back-end that translates proofs expressed in a theory encoded in Dedukti to other systems such as Coq, Lean or HOL 2) A front-end that prints these proofs in a "nice way" via a website. Using the website, the user can search for a definition or a theorem then, download the whole proof into the wanted system.

Currently, the available systems are: Coq, Matita, Lean, PVS and OpenTheory. The proofs comes from a logic called STTForall.

In the long run, more systems and more logic should be added.

RELEASE FUNCTIONAL DESCRIPTION: This is the beta version of Logipedia. It implements the functionalities mentioned above.

- Contact: François Thiré
- URL: http://www.logipedia.science

6. New Results

6.1. Implementation of Dedukti

During his master internship with Frédéric Blanqui and Bruno Barras, Gabriel Hondet developed a new rewrite engine for Dedukti [22]. The algorithm used in the new rewriting engine is formalised and a correctness proof is provided. This algorithm is based on the pattern matching algorithm by Maranget and used in OCaml. It is extended to rewriting rules, λ terms and non linear patterns. Some interesting implementation details are evinced and then we compare the efficiency of the new engine to a naive matching algorithm and to the rewriting engine of Dedukti. The results show that our implementation handles large rewrite systems better that the naive algorithm, and is always better than Dedukti's.

During her internship with Frédéric Blanqui and Emilio Gallego, Houda Mouzoun developed a Dedukti plugin for the VSCode editor.

During his internship with Frédéric Blanqui and Valentin Blot, Jui-Hsuan Wu implemented a prototype algorithm for deciding whether a function defined by rewriting rules is injective or not [23], and also a new algorithm proposed by Frédéric Blanqui for checking that user-defined rewrite rules preserve typing.

Bruno Barras has developed a reduction machine implementing a strong call-by-need strategy for β-reduction and pattern-matching. Higher-order pattern-matching is not yet fully implemented. Regarding efficiency, an exponential speed-up can be observed compared to the current call-by-name implementation on a large class of examples, but a constant slow-down shows on examples where call-by-name is the optimal strategy. With Beniamino Accattoli, he has started studying the correctness of this machine, without pattern-matching. They proved that the machine correctly implements β-reduction, but have no result yet regarding the strategy or the time complexity.

6.2. Theory of $\lambda\Pi$-calculus modulo rewriting

Dependency pairs are a key concept at the core of modern automated termination provers for first-order term rewriting systems. In [14][15], Frédéric Blanqui, Guillaume Genestier and Olivier Hermant introduced an extension of this technique for a large class of dependently-typed higher-order rewriting systems. This improves previous results by Wahlstedt on the one hand and the first author on the other hand to strong normalization and non-orthogonal rewriting systems. This new result has been implemented in the termination-checker SizeChangeTool [17], which participated in the Termination Competition and is used by Dedukti.

During his internship with Frédéric Blanqui and Valentin Blot, Jui-Hsuan Wu designed an algorithm for deciding whether a function defined by rewriting rules is injective or not [23]. This allows to improve the unification algorithm used in Dedukti for inferring types and missing arguments.

The expressiveness of dependent type theory can be extended by identifying types modulo some additional computation rules. But, for preserving the decidability of type-checking or the logical consistency of the system, one must make sure that those user-defined rewriting rules preserve typing. Frédéric Blanqui has developed a new method to check that property using Knuth-Bendix completion. A prototype implementation by Jui-Hsuan Wu is available in Dedukti.

Confluence is a crucial property of rewriting. Gaspard Férey and Jean-Pierre Jouannaud formalized the higher-order rewriting relation on untyped terms implemented in Dedukti and studied various criteria to obtain confluence of higher-order rewrite systems considered together with beta. In particular Von Oostrom's decreasing diagrams technique is applied to multi-steps extensions of simple term rewriting to achieve confluence criteria based on the decidable computation of (orthogonal) higher-order critical pairs. This work assumes left-linearity of rules for now but current work aims at extending these techniques to prove confluence of non-left-linear rule restricted to subsets of terms [20].

Fran cois Thiré has worked on a criterion that would help proving metatheoretical results on Cumulative Type Systems, such as expansion postponment and the equivalence between typed and untyped presentations of conversion. This has been published andpresented at LFMTP'19 [19]

Frédéric Gilbert has written a preprint about the definition of proof certificates for predicative subtyping [21].

6.3. Proof reconstruction

Proof assistants often call automated theorem provers to prove subgoals. However, each prover has its own proof calculus and the proof traces that it produces often lack many details to build a complete proof. Hence these traces are hard to check and reuse in proof assistants. Dedukti is a proof checker whose proofs can be translated to various proof assistants: Coq, HOL, Lean, Matita, PVS. Yacine El Haddad, Guillaume Burel and Frédéric Blanqui implemented a tool Ektraskto [16] that extracts TPTP subproblems from a TSTP file and reconstructs complete proofs in Dedukti using automated provers able to generate Dedukti proofs like ZenonModulo or ArchSAT. This tool is generic: it assumes nothing about the proof calculus of the prover producing the trace, and it can use different provers to produce the Dedukti proof. We applied our tool on traces produced by automated theorem provers on the CNF problems of the TPTP library and we were able to reconstruct a proof for a large proportion of them, significantly increasing the number of Dedukti proofs that could be obtained for those problems.

Zenon Modulo and iProverModulo, two automated theorem provers that can produce Dedukti proofs, have been presented in an article accepted in the Journal of Automated Reasoning [12].

6.4. Translating proofs to Dedukti

Agda is a dependently-typed programming language developed at Chalmers University, Gothenburg, Sweden, for 20 years. Thanks to the propositions-as-types correspondence of Curry-Howard, Agda is often used as a proof-assistant. Guillaume Genestier developed with Jesper Cockx a prototypical translator from Agda to Dedukti, which supports well some of the mainly used features of Agda and translates hundreds of definitions of the standard libraries. This implementation led to new encodings of theories in Dedukti, regarding: Universe Polymorphism, Inductive and Record Types, Dependent Pattern Matching, eta convertibility. The implementation of this translator permits to improve both Agda and Dedukti. Indeed, we discovered some bugged (almost not used) functions in Agda and had to extend some existing functions to our purpose. On the Dedukti side, this implementation was the first usage of the newly implemented feature of rewriting modulo associativity and commutativity, which required some minor improvements. Furthermore, our translation of eta-expansion using a defined function led to an improvement in the verification of type preservation of rewriting rules in Dedukti.

Isabelle is a logical framework developed at Technical University of Munich and Cambridge University since the 90s. It implements several logics such as HOL and ZF and is used as part of large verification projects such as seL4 and Flyspeck. Gabriel Hondet developed with Makarius Wenzel (from Augsburg) an export from Isabelle propositions to Dedukti, which was later extended by Michael Färber and Makarius Wenzel to export proofs. This required substantial work on the Isabelle kernel to extend the reconstruction of proof terms based on the work of Stefan Berghofer. Our newly developed proof export allows for an independent verification of a substantial portion of the Isabelle/HOL standard library as well as for the integration of results proved in Isabelle into Logipedia.

6.5. Models of cubical type theory

Bruno Barras and Rehan Malak have developed further their Dedukti library of presheaves. Using this library, they have built a semi-simplicial model of System F.

6.6. A proof system for PCTL and CTL*

Gilles Dowek, Ying Jiang, and Wu Peng, have proposed a proof system for the probabilistic modal logic PCTL. A paper is in preparation.

Gilles Dowek, Ying Jiang, Wu Peng, and Wenhui Zhang have started to study a proof system for CTL*, that mixes constructive and classical aspects.

The article Towards Combining Model Checking and Proof Checking, of Ying Jiang, Jian Liu, Gilles Dowek, and Kailiang Ji, has been published in The Computer Journal [13].

6.7. System I

Gilles Dowek and Alejandro Díaz-Caro have defined a lambda-calculus, the system I, to represent the proofs of a variant minimal propositional logic where isomorphic propositions are identified. Their paper Proof Normalisation in a Logic Identifying Isomorphic Propositions, has been presented at the International Conference on Formal Structures for Computation and Deduction. A second paper The virtues of eta-expansion in System I, showing that the addition of eta-expansion to system I actually simplies the system has been submitted to publication.

6.8. Computing with global environments

The call-by-need evaluation strategy for the λ-calculus is an evaluation strategy that lazily evaluates arguments only if needed, and if so, shares computations across all places where it is needed. To implement this evaluation strategy, abstract machines require some form of global environment. While abstract machines usually lead to a better understanding of the flow of control during the execution, easing in particular the definition of continuation-passing style translations, the case of machines with global environments turns out to be much more subtle.

In collaboration with Hugo Herbelin, Étienne Miquey introduced F_{Υ}, a calculus featuring a data type for typed stores and a mechanism of explicit coercions witnessing store extensions. This calculus defines a generic target of typed continuationand-environment-passing style translations for several calculi with global environment: it is compatible with different evaluation strategy (call-by-need, call-by-name, call-by-value) and different type systems (simple types, system F). On the logical side, these translations broadly amounts to a Kripke forcing-like translation mixed with a negative translation (for the continuation-passing part).

6.9. Computional interpretation of the axiom scheme of comprehension

The axiom scheme of comprehension is the cornerstone of second-order arithmetic, a logical theory in which most of mathematics can be formalized. Historically, comprehension was obtained from the negative translation of the axiom of choice, this axiom being interpreted by bar recursion. This led to cluttered and inefficient interpretations of second-order arithmetic.

Valentin Blot simplified this interpretation by proving that the axiom scheme of comprehension has a direct computational interpretation through a variant of bar recursion called update recursion. This new interpretation leads to a more efficient computational interpretation of proofs in second-order arithmetic, and paves the way for a convergence of the two existing interpretations: bar recursion and System F.

6.10. Alignment of logical connectives

Émilie Grienenberger and Gilles Dowek have studied in practice the alignment of logical connectives between proofs systems, a first step towards concept alignement, by the export of the HOL Light standard library using axiomatized connectives to Dedukti. More theoretically, an ecumenical system–where classical and intuitionistic logics coexist–was introduced to act as an exchange platform between proof systems.

6.11. Quantum Computing

The article Two linearities for quantum computing in the lambda calculus, of Alejandro Díaz-Caro, Gilles Dowek, and Juan Pablo Rinaldi, first published in the proceedings of Theory and Practice of Natural Computing 2017, has been published in the journal Biosystems.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Valentin Blot obtained with Chantal Keller funding for a 4-year project involving a PhD student, a research engineer (2 years) and a post-doctoral researcher (2 years). This funding is part of the Inria - Nomadic labs partnership for Tezos blockchain.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Valentin Blot obtained funding for hiring Étienne Miquey as a post-doctoral researcher from Île-de-France region's DIM-RFSI (Domaine d'Intérêt Majeur - Réseau Francilien en Sciences Informatiques).

8.2. National Initiatives

The ANR PROGRAMme is an ANR for junior researcher Liesbeth Demol (CNRS, UMR 8163 STL, University Lille 3) to which G. Dowek participates. The subject is: "What is a program? Historical and Philosophical perspectives". This project aims at developing the first coherent analysis and pluralistic understanding of "program" and its implications to theory and practice.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

Frédéric Blanqui cooperates with various researchers in Japan: Makato Hamana (Gunma University), Yoji Akama (Tohoku University) and Kentaro Kikuchi (Tohoku University).

8.4. International Research Visitors

8.4.1. Visits to International Teams

8.4.1.1. Research Stays Abroad

Gilles Dowek has spent two weeks at the Institute of Software in Beijing where he has worked with Ying Jiang, Wu Peng, and Wenhui Zhang.

Gilles Dowek has spent two weeks at the University of Buenos Aires where he has worked with Alejandro Díaz-Caro.

Frédéric Blanqui has been invited for two weeks in Japan by Yoji Akama (Tohoku University) and Makato Hamana (Gunma University).

As a "Short Term Scientific Mission" financed by COST Action EUTypes, Guillaume Genestier spent five weeks in Chalmers University, Gothenburg, Sweden, to cooperate with Jesper Cockx and Andreas Abel on the translation between the proof assistant Agda and Dedukti.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

Deducteam has organized the kick off meeting of Logipedia on 21-23 January 2019 in Cachan.

Frédéric Blanqui and Olivier Hermant organized the 11th International School on Rewriting (ISR) in Paris on 1-6 July 2019. The school offered to 37 participants from 8 countries, 69 hours of lectures in two parallel sessions, given by 15 lecturers from 6 countries.

Valentin Blot co-organized the Facets of realizability workshop on 1-3 July 2019, bringing together 22 researchers from 6 countries.

Frédéric Blanqui is Workshop Chair of the ACM/IEEE Symposium on Logic in Computer Science (LICS).

9.1.2. Scientific Events: Selection

9.1.2.1. Member of the Conference Program Committees

Gilles Dowek has been a member of the program committee of the Nasa Formal Methods Symposium (NFM'19) and Interactive Theorem Proving (ITP'19).

Frédéric Blanqui has been member of the program committee of the 16th International Colloquium on Theoretical Aspects of Computing (ICTAC'19), the 6th Workshop on Proof eXchange for Theorem Proving (PxTP'19), and of Large Mathematical Libraries (LML'19).

Valentin Blot was a member of the program committee for the french conference "Journées Francophones des Langages Applicatifs" 2020.

9.1.2.2. Reviewer

Guillaume Burel reviewed articles submitted to PxTP'19 and LOPSTR'19.

Valentin Blot reviewed articles submitted to LICS'19.

Bruno Barras reviewed an article submitted to CSL'20.

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing Activities

Guillaume Burel reviewed an article for Logical Methods in Computer Science.

Bruno Barras reviewed articles submitted to the Annals of Mathematics and Artificial Intelligence and Mathematical Structures in Computer Science

9.1.4. Invited Talks

Gilles Dowek has been invited to the meeting Big proofs in Edinburgh, where he has given a talk on Logipedia.

Gilles Dowek has participated to the meeting of the Proof society in Swansea where he has given a talk on Logipedia.

Gilles Dowek has given talk on Logipedia in Lille, Strasbourg, Beijing, Rio de Janeiro, and in the virtual "Laboratoire International de Recherche en Informatique et Mathématiques Appliquées".

Frédéric Blanqui gave talks at Gunma University (Japan) and Tohoku University (Japan).

Guillaume Genestier gave a seminar at the logic group of Chalmers University (Sweden).

Valentin Blot gave an invited talk at a realizability workshop on 18-20 September 2019 in Marseille, and a talk at the IRIF semantics working group.

9.1.5. Leadership within the Scientific Community

Frédéric Blanqui is member of the Steering Committees of LICS, TYPES and ISR.

9.1.6. Scientific Expertise

Bruno Barras reviewed a research proposal submitted to the Dutch National Research Organization (NWO).

9.1.7. Research Administration

Frédéric Blanqui is co-director of the pole 4 of the doctoral school STIC of the University Paris-Saclay.

Frédéric Blanqui is member of the committee of the doctoral school of the ENS Paris-Saclay.

Frédéric Blanqui is in charge of following PhD students at LSV.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Master: Frédéric Blanqui, formal languages, 21h, M1, ENSIIE, France
- Master: Frédéric Blanqui, rewriting theory, 14h, M1, ENS Paris-Saclay, France
- Master: Frédéric Blanqui, λ-calculus and theories in first-order logic, 18h, M1/M2, ENS Paris-Saclay, France
- Master: Bruno Barras, Proof Assistants, 12h, M2, MPRI, France
- Master: Gabriel Hondet, rewriting theory TD, 14h, M!, ENS Paris-Saclay, France
- Licence: Gabriel Hondet, operating systems and architecture TD, 22.5h, L3, ENS Paris-Saclay, France
- Licence: Gabriel Hondet, programming TD, 12h, L3, ENS Paris-Saclay, France
- License: Gaspard Férey, Théorie des Langages, 44h, L3, EISTI
- IUT: Yacine El Haddad, Programmation Web Côté Serveur, 62h, IUT of Orsay

9.2.2. Supervision

- PhD in progress: Guillaume Genestier, termination in $\lambda\Pi$-calculus modulo theory, 01/10/17, Frédéric Blanqui and Olivier Hermant.
- PhD in progress: Mohamed Yacine El Haddad, using automated provers in proof assistants, 05/01/18, Frédéric Blanqui and Guillaume Burel.
- PhD in progress: Gabriel Hondet, translating PVS proofs to Dedukti, 01/10/19, Frédéric Blanqui and Gilles Dowek.

9.2.3. Juries

Bruno Barras and Valentin Blot have participated to the PhD jury of Youssef El Bakouny from Université St-Joseph, Beirut (Lebanon).

9.3. Popularization

9.3.1. Articles and contents

Articles

- The synthesis article Algorithmes et modèles : l'histoire d'une convergence, written a decade ago has been published in Pierre Mounoud, Leçons de mathématiques d'aujourd'hui : Volume 5 (Cassini,2019).

Books

- Gilles Dowek, Ce dont on ne peut parler, il faut l'écrire : Langues et langages (Le Pommier, 2019)

Manuscripts

- Gilles Dowek, How the physical Church-Turing thesis changed the concept of machine, manuscript, 2019.
- Gilles Dowek, Two consequences of the hypothesis that we are within the world, manuscript, 2019.
- Gilles Dowek, Instinct, language, and artificial intelligence, manuscript, 2019.

9.3.2. Interventions

Gilles Dowek has been invited to the PROGRAMme workshop "Machines" where he has presented a paper How the physical Church-Turing thesis changed the concept of machine.

Gilles Dowek has been invited to the Conference on Robotics, AI, and Humanity, Science, Ethics, and Policy organized by the Pontifical Academy of Sciences and the Pontifical Academy of Social Sciences where he has presented a paper Instinct, language, and artificial intelligence.

Gilles Dowek has been invited at the meeting Le hasard, le calcul et la vie, in Cerisy, where he has given a talk Un Chaos discret.

Gilles Dowek has been invited to the meeting Experiencing reality directly in Jerusalem where he has presented a paper Two consequences of the hypothesis that we are within the world.

10. Bibliography

Major publications by the team in recent years

[1] A. ASSAF, G. BUREL, R. CAUDERLIER, D. DELAHAYE, G. DOWEK, C. DUBOIS, F. GILBERT, P. HALMAGRAND, O. HERMANT, R. SAILLARD.*Expressing theories in the λΠ-calculus modulo theory and in the Dedukti system*, in "22nd International Conference on Types for Proofs and Programs, TYPES 2016", Novi SAd, Serbia, May 2016, https://hal-mines-paristech.archives-ouvertes.fr/hal-01441751

[2] B. BARRAS, T. COQUAND, S. HUBER.*A generalization of the Takeuti-Gandy interpretation*, in "Mathematical Structures in Computer Science", 2015, vol. 25, n° 5, p. 1071–1099, https://doi.org/10.1017/S0960129514000504

[3] F. BLANQUI.*Definitions by rewriting in the Calculus of Constructions*, in "Mathematical Structures in Computer Science", 2005, vol. 15, n° 1, p. 37-92 [*DOI :* 10.1017/S0960129504004426], http://hal.inria.fr/inria-00105648/en/

[4] F. BLANQUI, J.-P. JOUANNAUD, A. RUBIO.*The Computability Path Ordering*, in "Logical Methods in Computer Science", October 2015 [*DOI :* 10.2168/LMCS-11(4:3)2015], https://hal.inria.fr/hal-01163091

[5] V. BLOT.*An interpretation of system F through bar recursion*, in "32nd ACM/IEEE Symposium on Logic in Computer Science", IEEE, 2017

[6] G. BUREL, G. BURY, R. CAUDERLIER, D. DELAHAYE, P. HALMAGRAND, O. HERMANT.*First-Order Automated Reasoning with Theories: When Deduction Modulo Theory Meets Practice*, in "Journal of Automated Reasoning", 2019 [*DOI :* 10.1007/S10817-019-09533-Z], https://hal.archives-ouvertes.fr/hal-02305831

[7] D. COUSINEAU, G. DOWEK.*Embedding Pure Type Systems in the λΠ-calculus modulo*, in "Typed lambda calculi and applications", S. RONCHI DELLA ROCCA (editor), Lecture Notes in Computer Science, Springer-Verlag, 2007, vol. 4583, p. 102-117

[8] G. DOWEK, T. HARDIN, C. KIRCHNER.*Theorem proving modulo*, in "Journal of Automated Reasoning", 2003, vol. 31, p. 33-73

[9] O. HERMANT.*Resolution is Cut-Free*, in "Journal of Automated Reasoning", March 2010, vol. 44, n° 3, p. 245-276

[10] M. JACQUEL, K. BERKANI, D. DELAHAYE, C. DUBOIS.*Verifying B Proof Rules using Deep Embedding and Automated Theorem Proving*, in "Software and Systems Modeling (SoSyM)", June 2013

[11] M. JACQUEL, K. BERKANI, D. DELAHAYE, C. DUBOIS.*Tableaux Modulo Theories Using Superdeduction*, in "Global Journal of Advanced Software Engineering (GJASE)", December 2014, vol. 1, p. 1-13 [*DOI :* 10.1007/978-3-642-31365-3_26], https://hal.archives-ouvertes.fr/hal-01099338

Publications of the year

Articles in International Peer-Reviewed Journal

[12] G. BUREL, G. BURY, R. CAUDERLIER, D. DELAHAYE, P. HALMAGRAND, O. HERMANT.*First-Order Automated Reasoning with Theories: When Deduction Modulo Theory Meets Practice*, in "Journal of Automated Reasoning", 2019, forthcoming [*DOI :* 10.1007/s10817-019-09533-z], https://hal.archives-ouvertes.fr/hal-02305831

[13] Y. JIANG, J. LIU, G. DOWEK, K. JI.*Towards Combining Model Checking and Proof Checking*, in "The Computer Journal", 2019, forthcoming [*DOI :* 10.1093/COMJNL/BXY112], https://hal.inria.fr/hal-01970274

International Conferences with Proceedings

[14] F. BLANQUI, G. GENESTIER, O. HERMANT.*Dependency Pairs Termination in Dependent Type Theory Modulo Rewriting*, in "FSCD 2019 - 4th International Conference on Formal Structures for Computation and Deduction", Dortmund, Germany, June 2019, https://arxiv.org/abs/1906.11649 [*DOI :* 10.4230/LIPIcs.FSCD.2019.9], https://hal.inria.fr/hal-01943941

[15] F. BLANQUI, G. GENESTIER, O. HERMANT.*Dependency Pairs Termination in Dependent Type Theory Modulo Rewriting*, in "TYPES 2019 - 25th International Conference on Types for Proofs and Programs", Oslo, Norway, M. BEZEM, N. VAN DER WEIDE (editors), June 2019, p. 30-31, https://hal.archives-ouvertes.fr/hal-02442484

[16] M. Y. EL HADDAD, G. BUREL, F. BLANQUI.*Ekstrakto A tool to reconstruct Dedukti proofs from TSTP files (extended abstract)*, in "PxTP 2019 - Sixth Workshop on Proof eXchange for Theorem Proving", Natal, Brazil, Electronic Proceedings in Theoretical Computer Science, August 2019, vol. 301, p. 27-35 [*DOI :* 10.4204/EPTCS.301.5], https://hal.inria.fr/hal-02200548

[17] G. GENESTIER.*SizeChangeTool: A Termination Checker for Rewriting Dependent Types*, in "HOR 2019 - 10th International Workshop on Higher-Order Rewriting", Dortmund, Germany, M. AYALA-RINCÓN, S. GHILEZAN, J. G. SIMONSEN (editors), Joint Proceedings of HOR 2019 and IWC 2019, June 2019, p. 14-19, https://hal.archives-ouvertes.fr/hal-02442465

National Conferences with Proceeding

[18] S. COLIN, R. LEPIGRE, G. SCHERER.*Unboxing Mutually Recursive Type Definitions in OCaml*, in "JFLA 2019 - 30 èmes journées francophones des langages applicatifs", Les Rousses, France, January 2019, https://arxiv.org/abs/1811.02300 , https://hal.inria.fr/hal-01929508

Conferences without Proceedings

[19] F. THIRÉ.*Cumulative Types Systems and Levels*, in "LFMTP 2019 - Logical Frameworks and Meta-Languages: Theory and Practice", Vancouver, Canada, June 2019, https://hal.archives-ouvertes.fr/hal-02150179

Other Publications

[20] G. FEREY, J.-P. JOUANNAUD.*Confluence in (Un)Typed Higher-Order Theories by means of Critical Pairs*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02096540

[21] F. GILBERT.*Verifiable certificates for predicate subtyping*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01977585

[22] G. HONDET.*Efficient rewriting using decision trees*, ENAC ; Univesité Toulouse III- Paul Sabatier, August 2019, https://hal.inria.fr/hal-02317471

[23] J.-H. WU.*Checking the type safety of rewrite rules in the $\lambda\Pi$-calculus modulo rewriting*, Ecole Normale Supérieure, September 2019, https://hal.inria.fr/hal-02288720

References in notes

[24] M. BOESPFLUG.*Conception d'un noyau de vérification de preuves pour le $\lambda\Pi$-calcul modulo*, École Polytechnique, 2011

[25] G. DOWEK.*A Theory Independent Curry-de Bruijn-howard Correspondence*, in "Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming - Volume Part II", Berlin, Heidelberg, ICALP'12, Springer-Verlag, 2012, p. 13–15, http://dx.doi.org/10.1007/978-3-642-31585-5_2

Project-Team DEFI

Shape reconstruction and identification

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

IN PARTNERSHIP WITH:
Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Numerical schemes and simulations

Table of contents

Project-Team DEFI

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.4. - Multiscale modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.3. - Probabilistic methods
A6.2.6. - Optimization
A6.2.7. - High performance computing
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A6.3.5. - Uncertainty Quantification

Other Research Topics and Application Domains:

B1.2.1. - Understanding and simulation of the brain and the nervous system
B1.2.3. - Computational neurosciences
B2.6.1. - Brain imaging
B3.2. - Climate and meteorology
B3.3.4. - Atmosphere

1. Team, Visitors, External Collaborators

Research Scientists

Houssem Haddar [Team leader, Inria, Senior Researcher]
Marcella Bonazzoli [Inria, Researcher]
Lucas Chesnel [Inria, Researcher]
Pietro-Marco Congedo [Inria, Researcher, HDR]
Jing-Rebecca Li [Inria, Researcher, HDR]

Faculty Member

Grégoire Allaire [École polytechnique, Professor, HDR]

External Collaborators

Lorenzo Audibert [EDF]
Olivier Le Maitre [CNRS]

Technical Staff

Nassim Auguste Razaaly Jamal [Inria, Engineer, from Aug 2019]
François Sanson [Inria, Engineer, from Aug 2019]

PhD Students

Julien Carlier [École polytechnique, PhD Student]
Chengran Fang [Inria, PhD Student, from Oct 2019]
Joao Miguel Felicio Dos Reis [Inria, PhD Student]

Hugo Girardon [EDF, PhD Student, granted by CIFRE]
Giulio Gori [Inria, PhD Student]
Jingjing Hao [Inria, PhD Student, until Jun 2019]
Nicolas Leoni [CEA, PhD Student, from Feb 2019]
Dorian Lereverend [École polytechnique, PhD Student, from Oct 2019]
Kevish Napal [École polytechnique, PhD Student]
Nassim Auguste Razaaly Jamal [Inria, PhD Student, from Feb 2019 until Jul 2019]
Mickael Rivier [Ecole Polytechnique, PhD Student, from Apr 2019, granted by CIFRE]
François Sanson [Inria, PhD Student, from Feb 2019 until Jul 2019]
Ana Isabel Del Val Benitez [VKI, Brussels]

Post-Doctoral Fellows

Marc Bakry [École polytechnique, Post-Doctoral Fellow]
Xiaoli Liu [Inria, Post-Doctoral Fellow]
Imen Mekkaoui [Inria, Post-Doctoral Fellow, until Jul 2019]

Visiting Scientists

Theo Bourdais [École polytechnique, from Sep 2019]
Nolan Burban [École polytechnique, from Sep 2019]
Fioralba Cakoni [Rutgers University, Nov 2019]
David Colton [University of Delaware, Nov 2019]
Marwa Kchaou [ENIT, from Jun 2019 until Jul 2019]
Mohamed Kamel Riahi [Khalifa University, May 2019]
Alessandro Romei [until Apr 2019]

Administrative Assistants

Natalia Alves [Inria, Administrative Assistant, until Jan 2019]
Marie Enee [Inria, Administrative Assistant, from Mar 2019]

2. Overall Objectives

2.1. Overall Objectives

The research activity of our team is dedicated to the design, analysis and implementation of efficient numerical methods to solve inverse and shape/topological optimization problems, eventually including system uncertainties, in connection with acoustics, electromagnetism, elastodynamics, diffusion, and fluid mechanics.

Sought practical applications include radar and sonar applications, bio-medical imaging techniques, non-destructive testing, structural design, composite materials, diffusion magnetic resonance imaging, fluid-driven applications in aerospace/energy fields.

Roughly speaking, the model problem consists in determining information on, or optimizing the geometry (topology) and the physical properties of unknown targets from given constraints or measurements, for instance, measurements of diffracted waves or induced magnetic fields. Moreover, system uncertainties can ben systematically taken into account to provide a measure of confidence of the numerical prediction.

In general this kind of problems is non-linear. The inverse ones are also severely ill-posed and therefore require special attention from regularization point of view, and non-trivial adaptations of classical optimization methods.

Our scientific research interests are the following:

- Theoretical understanding and analysis of the forward and inverse mathematical models, including in particular the development of simplified models for adequate asymptotic configurations.
- The design of efficient numerical optimization/inversion methods which are quick and robust with respect to noise. Special attention will be paid to algorithms capable of treating large scale problems (e.g. 3-D problems) and/or suited for real-time imaging.
- Propose new methods and develop advanced tools to perform uncertainty quantification for optimization/inversion.
- Development of prototype softwares for specific applications or tutorial toolboxes.

We were particularly interested in the development of the following themes

- Qualitative and quantitative methods for inverse scattering problems
- Topological optimization methods
- Forward and inverse models for Diffusion MRI
- Forward/Backward uncertainty quantification methods for optimization/inversion problems in the context of expensive computer codes.

3. Research Program

3.1. Research Program

The research activity of our team is dedicated to the design, analysis and implementation of efficient numerical methods to solve inverse and shape/topological optimization problems, eventually including system uncertainties, in connection with wave imaging, structural design, non-destructive testing and medical imaging modalities. We are particularly interested in the development of fast methods that are suited for real-time applications and/or large scale problems. These goals require to work on both the physical and the mathematical models involved and indeed a solid expertise in related numerical algorithms. A part of the research activity is also devoted to take into account system uncertainties in the solving of inverse/optimization problems. At the interface of physics, mathematics, and computer science, Uncertainty Quantification (UQ) focuses on the development of frameworks and methods to characterize uncertainties in predictive computations. Uncertainties and errors arise at different stages of the numerical simulation. First, errors are introduced due to the physical simplifications in the mathematical modeling of the system investigated; other errors come from the numerical resolution of the mathematical model, due in particular to finite discretization and computations with finite accuracy and tolerance; finally, errors are due a limited knowledge of input quantities (parameters) appearing in the definition of the numerical model being solved.

This section intends to give a general overview of our research interests and themes. We choose to present them through the specific academic example of inverse scattering problems (from inhomogeneities), which is representative of foreseen developments on both inversion and (topological) optimization methods. The practical problem would be to identify an inclusion from measurements of diffracted waves that result from the interaction of the sought inclusion with some (incident) waves sent into the probed medium. Typical applications include biomedical imaging where using micro-waves one would like to probe the presence of pathological cells, or imaging of urban infrastructures where using ground penetrating radars (GPR) one is interested in finding the location of buried facilities such as pipelines or waste deposits. This kind of applications requires in particular fast and reliable algorithms.

By "imaging" we refer to the inverse problem where the concern is only the location and the shape of the inclusion, while "identification" may also indicate getting informations on the inclusion physical parameters.

Both problems (imaging and identification) are non linear and ill-posed (lack of stability with respect to measurements errors if some careful constrains are not added). Moreover, the unique determination of the geometry or the coefficients is not guaranteed in general if sufficient measurements are not available. As an example, in the case of anisotropic inclusions, one can show that an appropriate set of data uniquely determine the geometry but not the material properties.

These theoretical considerations (uniqueness, stability) are not only important in understanding the mathematical properties of the inverse problem, but also guide the choice of appropriate numerical strategies (which information can be stably reconstructed) and also the design of appropriate regularization techniques. Moreover, uniqueness proofs are in general constructive proofs, i.e. they implicitly contain a numerical algorithm to solve the inverse problem, hence their importance for practical applications. The sampling methods introduced below are one example of such algorithms.

A large part of our research activity is dedicated to numerical methods applied to the first type of inverse problems, where only the geometrical information is sought. In its general setting the inverse problem is very challenging and no method can provide universally satisfying solution (respecting the balance cost-precision-stability). This is why in the majority of the practically employed algorithms, some simplification of the underlying mathematical model is used, according to the specific configuration of the imaging experiment. The most popular ones are geometric optics (the Kirchhoff approximation) for high frequencies and weak scattering (the Born approximation) for small contrasts or small obstacles. They actually give full satisfaction for a wide range of applications as attested by the large success of existing imaging devices (radar, sonar, ultrasound, X-ray tomography, etc.), that rely on one of these approximations.

In most cases, the used simplification result in a linearization of the inverse problem and therefore is usually valid only if the latter is weakly non-linear. The development of simplified models and the improvement of their efficiency is still a very active research area. With that perspective, we are particularly interested in deriving and studying higher order asymptotic models associated with small geometrical parameters such as: small obstacles, thin coatings, wires, periodic media, ... Higher order models usually introduce some non linearity in the inverse problem, but are in principle easier to handle from the numerical point of view than in the case of the exact model.

A larger part of our research activity is dedicated to algorithms that avoid the use of such approximations and that are efficient where classical approaches fail: i.e. roughly speaking when the non linearity of the inverse problem is sufficiently strong. This type of configuration is motivated by the applications mentioned below, and occurs as soon as the geometry of the unknown media generates non negligible multiple scattering effects (multiply-connected and closely spaces obstacles) or when the used frequency is in the so-called resonant region (wave-length comparable to the size of the sought medium). It is therefore much more difficult to deal with and requires new approaches. Our ideas to tackle this problem is mainly motivated and inspired by recent advances in shape and topological optimization methods and in so-called sampling methods.

Sampling methods are fast imaging solvers adapted to multi-static data (multiple receiver-transmitter pairs) at a fixed frequency. Even if they do not use any linearization the forward model, they rely on computing the solutions to a set of linear problems of small size, that can be performed in a completely parallel procedure. Our team has already a solid expertise in these methods applied to electromagnetic 3-D problems. The success of such approaches was their ability to provide a relatively quick algorithm for solving 3-D problems without any need for a priori knowledge on the physical parameters of the targets. These algorithms solve only the imaging problem, in the sense that only the geometrical information is provided.

Despite the large efforts already spent in the development of this type of methods, either from the algorithmic point of view or the theoretical one, numerous questions are still open. These attractive new algorithms also suffer from the lack of experimental validations, due to their relatively recent introduction. We also would like to invest on this side by developing collaborations with engineering research groups that have experimental facilities. From the practical point of view, the most potential limitation of sampling methods would be the need of a large amount of data to achieve a reasonable accuracy. On the other hand, optimization methods do not suffer from this constrain but they require good initial guess to ensure convergence and reduce the number of iterations. Therefore it seems natural to try to combine the two class of methods in order to calibrate the balance between cost and precision.

Among various shape optimization methods, the Level Set method seems to be particularly suited for such a coupling. First, because it shares similar mechanism as sampling methods: the geometry is captured as a level set of an “indicator function” computed on a cartesian grid. Second, because the two methods do not require any a priori knowledge on the topology of the sought geometry. Beyond the choice of a particular method, the main question would be to define in which way the coupling can be achieved. Obvious strategies consist in using one method to pre-process (initialization) or post-process (find the level set) the other. But one can also think of more elaborate ones, where for instance a sampling method can be used to optimize the choice of the incident wave at each iteration step. The latter point is closely related to the design of so called “focusing incident waves” (which are for instance the basis of applications of the time-reversal principle). In the frequency regime, these incident waves can be constructed from the eigenvalue decomposition of the data operator used by sampling methods. The theoretical and numerical investigations of these aspects are still not completely understood for electromagnetic or elastodynamic problems.

Other topological optimization methods, like the homogenization method or the topological gradient method, can also be used, each one provides particular advantages in specific configurations. It is evident that the development of these methods is very suited to inverse problems and provide substantial advantage compared to classical shape optimization methods based on boundary variation. Their applications to inverse problems has not been fully investigated. The efficiency of these optimization methods can also be increased for adequate asymptotic configurations. For instance small amplitude homogenization method can be used as an efficient relaxation method for the inverse problem in the presence of small contrasts. On the other hand, the topological gradient method has shown to perform well in localizing small inclusions with only one iteration.

A broader perspective would be the extension of the above mentioned techniques to time-dependent cases. Taking into account data in time domain is important for many practical applications, such as imaging in cluttered media, the design of absorbing coatings or also crash worthiness in the case of structural design.

For the identification problem, one would like to also have information on the physical properties of the targets. Of course optimization methods is a tool of choice for these problems. However, in some applications only a qualitative information is needed and obtaining it in a cheaper way can be performed using asymptotic theories combined with sampling methods. We also refer here to the use of so called transmission eigenvalues as qualitative indicators for non destructive testing of dielectrics.

We are also interested in parameter identification problems arising in diffusion-type problems. Our research here is mostly motivated by applications to the imaging of biological tissues with the technique of Diffusion Magnetic Resonance Imaging (DMRI). Roughly speaking DMRI gives a measure of the average distance travelled by water molecules in a certain medium and can give useful information on cellular structure and structural change when the medium is biological tissue. In particular, we would like to infer from DMRI measurements changes in the cellular volume fraction occurring upon various physiological or pathological conditions as well as the average cell size in the case of tumor imaging. The main challenges here are 1) correctly model measured signals using diffusive-type time-dependent PDEs 2) numerically handle the complexity of the tissues 3) use the first two to identify physically relevant parameters from measurements. For the last point we are particularly interested in constructing reduced models of the multiple-compartment Bloch-Torrey partial differential equation using homogenization methods.

The Team devotes a large effort focused on the formulation, implementation and validation of numerical methods for using scientific computing to drive experiments and available data (coming from models, simulation and experiments) by taking into account the system uncertainty. The team is also invested in exploiting the intimate relationship between optimisation and UQ to make Optimisation Under Uncertainty (OUU) tractable. A part of these activities is declined to the simulation of high-fidelity models for fluids, in three main fields, aerospace, energy and environment.

The Team is working on developing original UQ representations and algorithms to deal with complex and large scale models, having high dimensional input parameters with complexes influences. We are organizing our core research activities along different methodological UQ developments related to the challenges discussed above. Obviously, some efforts are shared by different initiatives or projects, and some of them include the continuous improvement of the non-intrusive methods constituting our software libraries. These actions are not detailed in the following, to focus the presentation on more innovative aspects, but we mentioned nonetheless the continuous developments and incorporation in our libraries of advanced sparse grid methods, sparsity promoting strategies and low rank methods.

An effort is dedicated to the efficient construction of surrogate models that are central in both forward and backward UQ problems, aiming at large-scale simulations relevant to engineering applications, with high dimensional input parameters.

Sensitivity analyses and other forward UQ problems (e.g., estimation of failure probabilities, rare events,. . .) depends on the input uncertainty model. Most often, for convenience or because of the lack of data, the independence of the uncertain inputs is assumed. In the Team, we are investigating approaches dedicated to a) the construction of uncertainty models that integrate the available information and expert knowledge(s) in a consistent and objective fashion. To this end, several mathematical frameworks are already available, e.g the maximum entropy principle, likelihood maximization and moment matching methods, but their application to real engineering problems remains scarce and their systematic use raises multiple challenges, both to construct the uncertainty model and to solve the related UQ problems (forward and backward). Because of the importance of the available data and expertise to build the model, the contributions of the Team in these areas depend on the needs and demands of end-users and industrial partners.

To mitigate computational complexity, the Team is exploring multi-fidelity approaches in the context of expensive simulations. We combine predictions of models with different levels of discretizations and physical simplifications to construct, at a controlled cost, reliable surrogate models of simulation outputs or directly objective functions and possibly constraints, to enable the resolution of robust optimization and stochastic inverse problems. Again, one difficulty to be addressed by the Team is the design of the computer experiments to obtain the best multi-fidelity model at the lowest cost (of for a prescribed computational budgets), with respect to the end use of the model. The last point is particularly challenging as it calls for accuracy for output values that are usually unknown a priori but must be estimated as the model construction proceeds.

4. Application Domains

4.1. Radar and GPR applications

Conventional radar imaging techniques (ISAR, GPR, etc.) use backscattering data to image targets. The commonly used inversion algorithms are mainly based on the use of weak scattering approximations such as the Born or Kirchhoff approximation leading to very simple linear models, but at the expense of ignoring multiple scattering and polarization effects. The success of such an approach is evident in the wide use of synthetic aperture radar techniques.

However, the use of backscattering data makes 3-D imaging a very challenging problem (it is not even well understood theoretically) and as pointed out by Brett Borden in the context of airborne radar: “In recent years it has become quite apparent that the problems associated with radar target identification efforts will not vanish with the development of more sensitive radar receivers or increased signal-to-noise levels. In addition it has (slowly) been realized that greater amounts of data - or even additional “kinds” of radar data, such as added polarization or greatly extended bandwidth - will all suffer from the same basic limitations affiliated with incorrect model assumptions. Moreover, in the face of these problems it is important to ask how (and if) the complications associated with radar based automatic target recognition can be surmounted.” This comment also applies to the more complex GPR problem.

Our research themes will incorporate the development, analysis and testing of several novel methods, such as sampling methods, level set methods or topological gradient methods, for ground penetrating radar application (imaging of urban infrastructures, landmines detection, underground waste deposits monitoring,) using multistatic data.

4.2. Biomedical imaging

Among emerging medical imaging techniques we are particularly interested in those using low to moderate frequency regimes. These include Microwave Tomography, Electrical Impedance Tomography and also the closely related Optical Tomography technique. They all have the advantage of being potentially safe and relatively cheap modalities and can also be used in complementarity with well established techniques such as X-ray computed tomography or Magnetic Resonance Imaging.

With these modalities tissues are differentiated and, consequentially can be imaged, based on differences in dielectric properties (some recent studies have proved that dielectric properties of biological tissues can be a strong indicator of the tissues functional and pathological conditions, for instance, tissue blood content, ischemia, infarction, hypoxia, malignancies, edema and others). The main challenge for these functionalities is to built a 3-D imaging algorithm capable of treating multi-static measurements to provide real-time images with highest (reasonably) expected resolutions and in a sufficiently robust way.

Another important biomedical application is brain imaging. We are for instance interested in the use of EEG and MEG techniques as complementary tools to MRI. They are applied for instance to localize epileptic centers or active zones (functional imaging). Here the problem is different and consists into performing passive imaging: the epileptic centers act as electrical sources and imaging is performed from measurements of induced currents. Incorporating the structure of the skull is primordial in improving the resolution of the imaging procedure. Doing this in a reasonably quick manner is still an active research area, and the use of asymptotic models would offer a promising solution to fix this issue.

4.3. Non destructive testing and parameter identification

One challenging problem in this vast area is the identification and imaging of defaults in anisotropic media. For instance this problem is of great importance in aeronautic constructions due to the growing use of composite materials. It also arises in applications linked with the evaluation of wood quality, like locating knots in timber in order to optimize timber-cutting in sawmills, or evaluating wood integrity before cutting trees. The anisotropy of the propagative media renders the analysis of diffracted waves more complex since one cannot only relies on the use of backscattered waves. Another difficulty comes from the fact that the micro-structure of the media is generally not well known a priori.

Our concern will be focused on the determination of qualitative information on the size of defaults and their physical properties rather than a complete imaging which for anisotropic media is in general impossible. For instance, in the case of homogeneous background, one can link the size of the inclusion and the index of refraction to the first eigenvalue of so-called interior transmission problem. These eigenvalues can be determined form the measured data and a rough localization of the default. Our goal is to extend this kind of idea to the cases where both the propagative media and the inclusion are anisotropic. The generalization to the case of cracks or screens has also to be investigated.

In the context of nuclear waste management many studies are conducted on the possibility of storing waste in a deep geological clay layer. To assess the reliability of such a storage without leakage it is necessary to have a precise knowledge of the porous media parameters (porosity, tortuosity, permeability, etc.). The large range of space and time scales involved in this process requires a high degree of precision as well as tight bounds on the uncertainties. Many physical experiments are conducted in situ which are designed for providing data for parameters identification. For example, the determination of the damaged zone (caused by excavation) around the repository area is of paramount importance since microcracks yield drastic changes in the permeability. Level set methods are a tool of choice for characterizing this damaged zone.

4.4. Diffusion MRI

In biological tissues, water is abundant and magnetic resonance imaging (MRI) exploits the magnetic property of the nucleus of the water proton. The imaging contrast (the variations in the grayscale in an image) in standard MRI can be from either proton density, T1 (spin-lattice) relaxation, or T2 (spin-spin) relaxation and the contrast in the image gives some information on the physiological properties of the biological tissue at different physical locations of the sample. The resolution of MRI is on the order of millimeters: the greyscale value shown in the imaging pixel represents the volume-averaged value taken over all the physical locations contained that pixel.

In diffusion MRI, the image contrast comes from a measure of the average distance the water molecules have moved (diffused) during a certain amount of time. The Pulsed Gradient Spin Echo (PGSE) sequence is a commonly used sequence of applied magnetic fields to encode the diffusion of water protons. The term 'pulsed' means that the magnetic fields are short in duration, an the term gradient means that the magnetic fields vary linearly in space along a particular direction. First, the water protons in tissue are labelled with nuclear spin at a precession frequency that varies as a function of the physical positions of the water molecules via the application of a pulsed (short in duration, lasting on the order of ten milliseconds) magnetic field. Because the precessing frequencies of the water molecules vary, the signal, which measures the aggregate phase of the water molecules, will be reduced due to phase cancellations. Some time (usually tens of milliseconds) after the first pulsed magnetic field, another pulsed magnetic field is applied to reverse the spins of the water molecules. The time between the applications of two pulsed magnetic fields is called the 'diffusion time'. If the water molecules have not moved during the diffusion time, the phase dispersion will be reversed, hence the signal loss will also be reversed, the signal is called refocused. However, if the molecules have moved during the diffusion time, the refocusing will be incomplete and the signal detected by the MRI scanner if weaker than if the water molecules have not moved. This lack of complete refocusing is called the signal attenuation and is the basis of the image contrast in DMRI. the pixels showning more signal attenuation is associated with further water displacement during the diffusion time, which may be linked to physiological factors, such as higher cell membrane permeability, larger cell sizes, higher extra-cellular volume fraction.

We model the nuclear magnetization of water protons in a sample due to diffusion-encoding magnetic fields by a multiple compartment Bloch-Torrey partial differential equation, which is a diffusive-type time-dependent PDE. The DMRI signal is the integral of the solution of the Bloch-Torrey PDE. In a homogeneous medium, the intrinsic diffusion coeffcient D will appear as the slope of the semi-log plot of the signal (in approporiate units). However, because during typical scanning times, 50-100ms, water molecules have had time to travel a diffusion distance which is long compared to the average size of the cells, the slope of the semi-log plot of the signal is in fact a measure of an 'effective' diffusion coefficient. In DMRI applications, this measured quantity is called the 'apparent diffusion coefficient' (ADC) and provides the most commonly used form the image contrast for DMRI. This ADC is closely related to the effective diffusion coefficient obtainable from mathematical homogenization theory.

4.5. Fluid flow applications

Specific actions are devoted to the problem of atmospheric reentry simulations. We focus on several aspects : i) on the development of innovative algorithms improving the prediction of hypersonic flows and including system uncertainties, ii) on the application of these methods to the atmospheric reentry of space vehicles for the control and the optimization of the trajectory, iii) on the debris reentry, which is of fundamental importance for NASA, CNES and ESA. Several works are already initiated with funding from CNES, Thales, and ASL. An ongoing activity concerns the design of the Thermal Protection System (TPS) that shields the spacecraft from aerothermal heating, generated by friction at the surface of the vehicle. The TPS is usually composed of different classes of materials, depending on the mission and the planned trajectory. One major issue is to model accurately the material response to ensure a safe design. High-fidelity material modeling for ablative materials has been developed by NASA, but a lot of work is still needed concerning the assessment of physical and modeling uncertainties during the design process. Our objective is to set up a predictive numerical tool to reliably estimate the response of ablative materials for different aerothermal conditions.

An important effort is dedicated to the simulation of fluids featuring complex thermodynamic behavior, in the context of two distinct projects: the VIPER project, funded by Aquitaine Region, and a project with CWI (Scientific Computing Group). Dense gases (DGs) are defined as single-phase vapors operating at temperatures and pressures conditions close to the saturation curve. The interest in studying complex dynamics of compressible dense gas flows comes from the potential technological advantages of using these fluids in energy conversion cycles, such as in Organic Rankine Cycles (ORCs) which used dense gases as energy converters for biomass fuels and low-grade heat from geothermal or industrial waste heat sources. Since these fluids feature large uncertainties in their estimated thermodynamic properties (critical properties, acentric factor, etc.), a meaningful numerical prediction of the performance must necessarily take into account these uncertainties. Other sources of uncertainties include, but are not limited to, the inlet boundary conditions which are often unknown in dense gases applications. Moreover, a robust optimization must also include the more generic uncertainty introduced by the machining tolerance in the construction of the turbine blades.

5. New Results

5.1. Qualitative and quantitative methods for inverse problems

5.1.1. Differential tomography of micromechanical evolution in elastic materials of unknown micro/macrostructure

H. haddar and F. Pourahmadian

Differential evolution indicators are introduced for 3D spatiotemporal imaging of micromechanical processes in complex materials where progressive variations due to manufacturing and/or aging are housed in a highly scattering background of a-priori unknown or uncertain structure. In this vein, a three-tier imaging platform is established where: (1) the domain is periodically (or continuously) subject to illumination and sensing in an arbitrary configuration; (2) sequential sets of measured data are deployed to distill segment-wise scattering signatures of the domain's internal structure through carefully constructed, non-iterative solutions to the scattering equation; and (3) the resulting solution sequence is then used to rigorously construct an imaging functional carrying appropriate invariance with respect to the unknown stationary components of the background e.g., pre-existing interstitial boundaries and bubbles. This gives birth to differential indicators that specifically recover the 3D support of micromechanical evolution within a network of unknown scatterers. The direct scattering problem is formulated in the frequency domain where the background is comprised of a random distribution of monolithic fragments. The constituents are connected via highly heterogeneous interfaces of unknown elasticity and dissipation which are subject to spatiotemporal evolution. The support of internal boundaries are sequentially illuminated by a set of incident waves and thusinduced scattered fields are captured over a generic observation surface. The performance of the proposed imaging indicator is illustrated through a set of numerical experiments for spatiotemporal reconstruction of progressive damage zones featuring randomly distributed cracks and bubbles [35].

5.1.2. Microwave tomographic imaging of cerebrovascular accidents by using high-performance computing

P.-H. Tournier, I. Aliferis, M. Bonazzoli, M. De Buhan, M. Darbas, V. Dolean, F. Hecht, P. Jolivet, I. El Kanfoud, C. Migliaccio, F. Nataf, C. Pichot, S. Semenov

The motivation of this work is the detection of cerebrovascular accidents by microwave tomographic imaging. This requires the solution of an inverse problem relying on a minimization algorithm (for example, gradient-based), where successive iterations consist in repeated solutions of a direct problem. The reconstruction algorithm is extremely computationally intensive and makes use of efficient parallel algorithms and high-performance computing. The feasibility of this type of imaging is conditioned on one hand by an accurate reconstruction of the material properties of the propagation medium and on the other hand by a considerable reduction in simulation time. Fulfilling these two requirements will enable a very rapid and accurate diagnosis. From the mathematical and numerical point of view, this means solving Maxwell's equations in time-harmonic regime by appropriate domain decomposition methods, which are naturally adapted to parallel architectures [20].

5.1.3. A Factorization Method for Inverse Time Domain Obstacles with Impedance Boundary Conditions

H. Haddar and X. Liu

We develop a factorization method to obtain explicit characterization of a (possibly non-convex) impedance scattering object from measurements of time-dependent causal scattered waves in the far field regime. In particular, we prove that far fields of solutions to the wave equation due to particularly modified incident waves, characterize the obstacle by a range criterion involving the square root of the time derivative of the corresponding far field operator. Our analysis makes essential use of a coercivity property of the solution of the initial boundary value problem for the wave equation in the Laplace domain. This forces us to consider this particular modification of the far field operator. The latter in fact, can be chosen arbitrarily close to the true far field operator given in terms of physical measurements. We provide validating numerical examples in 2D on synthetic data. The latter is generated using a FDTD solver with PML. An article on this topic is under preparation.

5.1.4. A robust Expectation-Maximization method for the interpretation of small angle scattering data on dense nanoparticle samples

M. Bakry, H. Haddar and O. Bunau

The Local Monodisperse Approximation (LMA) is a two-parameters model commonly employed for the retrieval of size distributions from the small angle scattering (SAS) patterns obtained on dense nanoparticle samples (e.g. dry powders and concentrated solutions). This work features an original, beyond state-of-the-art implementation of the LMA model resolution for the inverse scattering problem. Our method is based on the Expectation Maximization iterative algorithm and is free from any fine tuning of model parameters. The application of our method on SAS data acquired in laboratory conditions on dense nanoparticle samples is shown to provide very good results [2].

5.1.5. Factorization Method for Imaging a Local Perturbation in Inhomogeneous Periodic Layers from Far Field Measurements

H. Haddar and A. Konschin

We analyze the Factorization method to reconstruct the geometry of a local defect in a periodic absorbing layer using almost only incident plane waves at a fixed frequency. A crucial part of our analysis relies on the consideration of the range of a carefully designed far field operator, which characterizes the geometry of the defect. We further provide some validating numerical results in a two dimensional setting [9].

5.1.6. Shape reconstruction of deposits inside a steam generator using eddy current measurements

L. Audibert, H. Girardon and H. Haddar

Non-destructive testing is an essential tool to assess the safety of the facilities within nuclear plants. In particular, conductive deposits on U-tubes in steam generators constitute a major danger as they may block the cooling loop. To detect these deposits, eddy-current probes are introduced inside the U-tubes to generate currents and measuring back an impedance signal. Based on earlier work on this subject, we develop a shape optimization technique with regularized gradient descent to invert these measurements and recover the deposit shape. To deal with the unknown, and possibly complex, topological nature of the latter, we propose to model it using a level set function. The methodology is first validated on synthetic axisymmetric configurations and fast convergence in ensured by careful adaptation of the gradient steps and regularization parameters. We then consider a more realistic modeling that incorporates the support plate and the presence of imperfections on the tube interior section. We employ in particular an asymptotic model to take into account these imperfections and treat them as additional unknowns in our inverse problem. A multi-objective optimization strategy, based on the use of different operating frequencies, is then developed to solve this problem. Various numerical experimentation with synthetic data demonstrated the viability of our approach.

The approach is also successfully validated against experimental data. An article on this topic is under preparation.

5.1.7. On quasi-reversibility solutions to the Cauchy problem for the Laplace equation: regularity and error estimates

L. Bourgeois, L. Chesnel

We are interested in the classical ill-posed Cauchy problem for the Laplace equation. One method to approximate the solution associated with compatible data consists in considering a family of regularized well-posed problems depending on a small parameter $\varepsilon > 0$. In this context, in order to prove convergence of finite elements methods, it is necessary to get regularity results of the solutions to these regularized problems which hold uniformly in ε. In the present work, we obtain these results in smooth domains and in 2D polygonal geometries. In presence of corners, due the particular structure of the regularized problems, classical techniques *à la* Grisvard do not work and instead, we apply the Kondratiev approach. We describe the procedure in detail to keep track of the dependence in ε in all the estimates. The main originality of this study lies in the fact that the limit problem is ill-posed in any framework [4].

5.1.8. Data Completion Method For the Helmholtz Equation Via Surface Potentials for Partial Cauchy Data

M. Aussal, Y. Boukari and H. Haddar

We propose and study a data completion algorithm for recovering missing data from the knowledge of Cauchy data on parts of the same boundary. The algorithm is based on surface representation of the solution and is presented for the Helmholtz equation. This work is an extension of the data completion algorithm proposed by the two last authors where the case of data available of a closed boundary was studied. The proposed method is a direct inversion method robust with respect to noisy incompatible data. Classical regularization methods with discrepancy selection principles can be employed and automatically lead to a convergent schemes as the noise level goes to zero. We conduct 3D numerical investigations to validate our method on various synthetic examples [31].

5.2. Invisiblity and transmission eigenvalues

5.2.1. Inside-outside duality with artificial backgrounds

L. Audibert, L. Chesnel, H. Haddar

We use the inside-outside duality approach proposed by Kirsch-Lechleiter to identify transmission eigenvalues associated with artificial backgrounds. We prove that for well chosen artificial backgrounds, in particular for the ones with zero index of refraction at the inclusion location, one obtains a necessary and sufficient condition characterizing transmission eigenvalues via the spectrum of the modified far field operator. We also complement the existing literature with a convergence result for the invisible generalized incident field associated with the transmission eigenvalues [1].

5.2.2. Surface waves in a channel with thin tunnels and wells at the bottom: non-reflecting underwater topography

L. Chesnel, S.A. Nazarov, J. Taskinen

We consider the propagation of surface water waves in a straight planar channel perturbed at the bottom by several thin curved tunnels and wells. We propose a method to construct non reflecting underwater topographies of this type at an arbitrary prescribed wave number. To proceed, we compute asymptotic expansions of the diffraction solutions with respect to the small parameter of the geometry taking into account the existence of boundary layer phenomena. We establish error estimates to validate the expansions using advances techniques of weighted spaces with detached asymptotics. In the process, we show the absence of trapped surface waves for perturbations small enough. This analysis furnishes asymptotic formulas for the scattering matrix and we use them to determine underwater topographies which are non-reflecting. Theoretical and numerical examples are given [6]

5.2.3. Exact zero transmission during the Fano resonance phenomenon in non symmetric waveguides

L. Chesnel, S.A. Nazarov

We investigate a time-harmonic wave problem in a waveguide. We work at low frequency so that only one mode can propagate. It is known that the scattering matrix exhibits a rapid variation for real frequencies in a vicinity of a complex resonance located close to the real axis. This is the so-called Fano resonance phenomenon. And when the geometry presents certain properties of symmetry, there are two different real frequencies such that we have either $R = 0$ or $T = 0$, where R and T denote the reflection and transmission coefficients. In this work, we prove that without the assumption of symmetry of the geometry, quite surprisingly, there is always one real frequency for which we have $T = 0$. In this situation, all the energy sent in the waveguide is backscattered. However in general, we do not have $R = 0$ in the process. We provide numerical results to illustrate our theorems [33].

5.2.4. Homogenization of Maxwell's equations and related scalar problems with sign-changing coefficients

R. Bunoiu, L. Chesnel, K. Ramdani, M. Rihani

In this work, we are interested in the homogenization of time-harmonic Maxwell's equations in a composite medium with periodically distributed small inclusions of a negative material. Here a negative material is a material modelled by negative permittivity and permeability. Due to the sign-changing coefficients in the equations, it is not straightforward to obtain uniform energy estimates to apply the usual homogenization techniques. The goal of this work is to explain how to proceed in this context. The analysis of Maxwell's equations is based on a precise study of two associated scalar problems: one involving the sign-changing permittivity with Dirichlet boundary conditions, another involving the sign-changing permeability with Neumann boundary conditions. For both problems, we obtain a criterion on the physical parameters ensuring uniform invertibility of the corresponding operators as the size of the inclusions tends to zero. In the process, we explain the link existing with the so-called Neumann-Poincaré operator, complementing the existing literature on this topic. Then we use the results obtained for the scalar problems to derive uniform energy estimates for Maxwell's system. At this stage, an additional difficulty comes from the fact that Maxwell's equations are also sign-indefinite due to the term involving the frequency. To cope with it, we establish some sort of uniform compactness result [32].

5.3. Shape and topology optimization

5.3.1. Null space gradient flows for constrained optimization with applications to shape optimization

G. Allaire, F. Feppon and C. Dapogny

The purpose of this article is to introduce a gradient-flow algorithm for solving equality and inequality constrained optimization problems, which is particularly suited for shape optimization applications. We rely on a variant of the Ordinary Differential Equation (ODE) approach proposed by Yamashita for equality constrained problems: the search direction is a combination of a null space step and a range space step, aiming to decrease the value of the minimized objective function and the violation of the constraints, respectively. Our first contribution is to propose an extension of this ODE approach to optimization problems featuring both equality and inequality constraints. In the literature, a common practice consists in reducing inequality constraints to equality constraints by the introduction of additional slack variables. Here, we rather solve their local combinatorial character by computing the projection of the gradient of the objective function onto the cone of feasible directions. This is achieved by solving a dual quadratic programming subproblem whose size equals the number of active or violated constraints. The solution to this problem allows to identify the inequality constraints to which the optimization trajectory should remain tangent. Our second contribution is a formulation of our gradient flow in the context of—infinite-dimensional—Hilbert spaces, and of even more general optimization sets such as sets of shapes, as it occurs in shape optimization within the framework of Hadamard's boundary variation method. The cornerstone of this formulation is the classical operation of extension and regularization of shape derivatives. The numerical efficiency and ease of implementation of our algorithm are demonstrated on realistic shape optimization problems. An article on this topic is under preparation.

5.3.2. *A variational formulation for computing shape derivatives of geometric constraints along rays*

G. Allaire, F. Feppon and C. Dapogny

In the formulation of shape optimization problems, multiple geometric constraint functionals involve the signed distance function to the optimized shape Ω. The numerical evaluation of their shape derivatives requires to integrate some quantities along the normal rays to Ω, a task that is usually achieved thanks to the method of characteristics. The goal of the present paper is to propose an alternative, variational approach for this purpose. Our method amounts, in full generality, to compute integral quantities along the characteristic curves of a given velocity field without requiring the explicit knowledge of these curves on the spatial discretization; it rather relies on a variational problem which can be solved conveniently by the finite element method. The well-posedness of this problem is established thanks to a detailed analysis of weighted graph spaces of the advection operator $\beta \cdot \nabla$ associated to a $\mathcal{C}^1$ velocity fields β. One novelty of our approach is the ability to handle velocity fields with possibly unbounded divergence: we do not assume $\div(\beta) \in L^\infty$. Our working assumptions are fulfilled in the context of shape optimization of $\mathcal{C}^2$ domains Ω, where the velocity field $\beta = \nabla d_\Omega$ is an extension of the unit outward normal vector to the optimized shape. The efficiency of our variational method with respect to the direct integration of numerical quantities along rays is evaluated on several numerical examples. Classical albeit important implementation issues such as the calculation of a shape's curvature and the detection of its skeleton are discussed. Finally, we demonstrate the convenience and potential of our method when it comes to enforcing maximum and minimum thickness constraints in structural shape optimization. An article on this topic is under preparation.

5.3.3. *3-d topology optimization of modulated and oriented periodic microstructures by the homogenization method*

G. Allaire, P. Geoffroy-Donders and O. Pantz

This paper is motivated by the optimization of so-called lattice materials which are becoming increasingly popular in the context of additive manufacturing. Generalizing our previous work in 2-d we propose a method for topology optimization of structures made of periodically perforated material, where the microscopic periodic cell can be macroscopically modulated and oriented. This method is made of three steps. The first step amounts to compute the homogenized properties of an adequately chosen parametrized microstructure (here, a cubic lattice with varying bar thicknesses). The second step optimizes the homogenized formulation of the problem, which is a classical problem of parametric optimization. The third, and most delicate, step projects the optimal oriented microstructure at a desired length scale. Compared to the 2-d case where rotations are parametrized by a single angle, to which a conformality constraint can be applied, the 3-d case is more involved and requires new ingredients. In particular, the full rotation matrix is regularized (instead of just one angle in 2-d) and the projection map which deforms the square periodic lattice is computed component by component. Several numerical examples are presented for compliance minimization in 3-d. An article on this topic is under preparation.

5.4. Analysis of some wave problems

5.4.1. *On well-posedness of time-harmonic problems in an unbounded strip for a thin plate model*

L. Bourgeois, L. Chesnel, S. Fliss

We study the propagation of elastic waves in the time-harmonic regime in a waveguide which is unbounded in one direction and bounded in the two other (transverse) directions. We assume that the waveguide is thin in one of these transverse directions, which leads us to consider a Kirchhoff-Love plate model in a locally perturbed 2D strip. For time harmonic scattering problems in unbounded domains, well-posedness does not hold in a classical setting and it is necessary to prescribe the behaviour of the solution at infinity. This is challenging for the model that we consider and constitutes our main contribution. Two types of boundary conditions are considered: either the strip is simply supported or the strip is clamped. The two boundary conditions are treated with two different methods. For the simply supported problem, the analysis is based on a result of Hilbert basis in the transverse section. For the clamped problem, this property does not hold. Instead we adopt the Kondratiev's approach, based on the use of the Fourier transform in the unbounded direction, together with techniques of weighted Sobolev spaces with detached asymptotics. After introducing radiation conditions, the corresponding scattering problems are shown to be well-posed in the Fredholm sense. We also show that the solutions are the physical (outgoing) solutions in the sense of the limiting absorption principle. [5]

5.4.2. Domain decomposition preconditioning for the high-frequency time-harmonic Maxwell equations with absorption

M. Bonazzoli, V. Dolean, I. G. Graham, E. A. Spence and P.-H. Tournier

In this work we rigorously analyse preconditioners for the time-harmonic Maxwell equations with absorption, where the PDE is discretised using curl-conforming finite-element methods of fixed, arbitrary order and the preconditioner is constructed using additive Schwarz domain decomposition methods. The theory we developed shows that if the absorption is large enough, and if the subdomain and coarse mesh diameters and overlap are chosen appropriately, then the classical two-level overlapping additive Schwarz preconditioner (with PEC boundary conditions on the subdomains) performs optimally–in the sense that GMRES converges in a wavenumber-independent number of iterations–for the problem with absorption. An important feature of the theory is that it allows the coarse space to be built from low-order elements even if the PDE is discretised using high-order elements. It also shows that additive methods with minimal overlap can be robust. Several numerical experiments illustrate the theory and its dependence on various parameters. These experiments motivate some extensions of the preconditioners which have better robustness for problems with less absorption, including the propagative case. Finally, we illustrate the performance of these on two substantial applications; the first (a problem with absorption arising from medical imaging) shows the empirical robustness of the preconditioner against heterogeneity, and the second (scattering by a COBRA cavity) shows good scalability of the preconditioner with up to 3,000 processors [3].

5.4.3. Multi-Trace FEM-BEM formulation for acoustic scattering by composite objects

M. Bonazzoli, X. Claeys

This work is about the scattering of an acoustic wave by an object composed of piecewise homogenous parts and an arbitrarily heterogeneous part. We propose and analyze a formulation that couples, adopting a Costabel-type approach, boundary integral equations for the homogeneous subdomains with domain variational formulations for the heterogenous subdomain. This is an extension of Costabel FEM-BEM coupling to a multi-domain configuration, with junctions points allowed, i.e. points where three or more subdomains abut. Usually just the exterior unbounded subdomain is treated with the BEM; here we wish to exploit the BEM whenever it is applicable, that is for all the homogenous parts of the scattering object, since it yields a reduction in the number of unknowns compared to the FEM. Our formulation is based on the multi-trace formalism for acoustic scattering by piecewise homogeneous objects; here we allow the wavenumber to vary arbitrarily in a part of the domain. We prove that the bilinear form associated with the proposed formulation satisfies a Gårding coercivity inequality, which ensures stability of the variational problem if it is uniquely solvable. We identify conditions for injectivity and construct modified versions immune to spurious resonances. An article on this topic is under preparation.

5.4.4. Domain decomposition preconditioners for non-self-adjoint or indefinite problems

M. Bonazzoli, X. Claeys, F. Nataf, P.-H. Tournier

The matrices arising from the finite element discretization of problems such as high-frequency Helmholtz, time-harmonic Maxwell or convection-diffusion equations are not self-adjoint or positive definite. For this reason, it is difficult to analyze the convergence of Schwarz domain decomposition preconditioners applied to these problems. Note also that the conjugate gradient method cannot be used, and the analysis of the spectrum of the preconditioned matrix is not sufficient for methods suited for general matrices such as GMRES. In order to apply Elman-type estimates for the convergence of GMRES we need to prove an upper bound on the norm of the preconditioned matrix, and a lower bound on the distance of its field of values from the origin. We generalize the theory for the Helmholtz equation developed for the SORAS (Symmetrized Optimized Restricted Additive Schwarz) preconditioner, and we identify a list of assumptions and estimates that are sufficient to prove the two bounds needed for the convergence analysis for a general linear system. As an illustration of this technique, we prove estimates for the heterogenous reaction-convection-diffusion equation. An article on this topic is under preparation.

5.5. Diffusion MRI

J.-R. Li, K. V. Nguyen and T. N. Tran

Diffusion Magnetic Resonance Imaging (DMRI) is a promising tool to obtain useful information on microscopic structure and has been extensively applied to biological tissues.

We obtained the following results.

- The Bloch-Torrey partial differential equation can be used to describe the evolution of the transverse magnetization of the imaged sample under the influence of diffusion-encoding magnetic field gradients inside the MRI scanner. The integral of the magnetization inside a voxel gives the simulated diffusion MRI signal. This work proposes a finite element discretization on manifolds in order to efficiently simulate the diffusion MRI signal in domains that have a thin layer or a thin tube geometrical structure. The variable thickness of the three-dimensional domains is included in the weak formulation established on the manifolds. We conducted a numerical study of the proposed approach by simulating the diffusion MRI signals from the extracellular space (a thin layer medium) and from neurons (a thin tube medium), comparing the results with the reference signals obtained using a standard three-dimensional finite element discretization. We show good agreements between the simulated signals using our proposed method and the reference signals for a wide range of diffusion MRI parameters. The approximation becomes better as the diffusion time increases. The method helps to significantly reduce the required simulation time, computational memory, and difficulties associated with mesh generation, thus opening the possibilities to simulating complicated structures at low cost for a better understanding of diffusion MRI in the brain [12].
- The nerve cells of the Aplysia are much larger than mammalian neurons. Using the Aplysia ganglia to study the relationship between the cellular structure and the diffusion MRI signal can potentially shed light on this relationship for more complex organisms. We measured the dMRI signal of chemically-fixed abdominal ganglia of the Aplysia at several diffusion times. At the diffusion times measured and observed at low b-values, the dMRI signal is mono-exponential and can be accurately represented by the parameter ADC (Apparent Diffusion Coefficient). We performed numerical simulations of water diffusion for the large cell neurons in the abdominal ganglia after creating geometrical configurations by segmenting high resolution T2-weighted (T2w) images to obtain the cell outline and then incorporating a manually generated nucleus. The results of the numerical simulations validate the claim that water diffusion in the large cell neurons is in the short diffusion time regime at our experimental diffusion times. Then, using the analytical short time approximation (STA) formula for the ADC, we showed that in order to explain the experimentally observed behavior, it is necessary to consider the nucleus and the cytoplasm as two separate diffusion compartments. By using a two compartment STA model, we were able to illustrate the effect of the highly irregular shape of the cell nucleus on the ADC [13].
- The complex transverse water proton magnetization subject to diffusion-encoding magnetic field

gradient pulses in a heterogeneous medium can be modeled by the multiple compartment Bloch-Torrey partial differential equation. Under the assumption of negligible water exchange between compartments, the time-dependent apparent diffusion coefficient can be directly computed from the solution of a diffusion equation subject to a time-dependent Neumann boundary condition. This work describes a publicly available MATLAB toolbox called SpinDoctor that can be used 1) to solve the Bloch-Torrey partial differential equation in order to simulate the diffusion magnetic resonance imaging signal; 2) to solve a diffusion partial differential equation to obtain directly the apparent diffusion coefficient; 3) to compare the simulated apparent diffusion coefficient with a short-time approximation formula. The partial differential equations are solved by $P1$ finite elements combined with built-in MATLAB routines for solving ordinary differential equations. The finite element mesh generation is performed using an external package called Tetgen. SpinDoctor provides built-in options of including 1) spherical cells with a nucleus; 2) cylindrical cells with a myelin layer; 3) an extra-cellular space enclosed either a) in a box or b) in a tight wrapping around the cells; 4) deformation of canonical cells by bending and twisting; 5) permeable membranes; Built-in diffusion-encoding pulse sequences include the Pulsed Gradient Spin Echo and the Oscillating Gradient Spin Echo. We describe in detail how to use the SpinDoctor toolbox. We validate SpinDoctor simulations using reference signals computed by the Matrix Formalism method. We compare the accuracy and computational time of SpinDoctor simulations with Monte-Carlo simulations and show significant speed-up of SpinDoctor over Monte-Carlo simulations in complex geometries. We also illustrate several extensions of SpinDoctor functionalities, including the incorporation of $T2$ relaxation, the simulation of non-standard diffusion-encoding sequences, as well as the use of externally generated geometrical meshes [10].

- The numerical simulation of the diffusion MRI signal arising from complex tissue micro-structures is helpful for understanding and interpreting imaging data as well as for designing and optimizing MRI sequences. The discretization of the Bloch-Torrey equation by finite elements is a more recently developed approach for this purpose, in contrast to random walk simulations, which has a longer history. While finite element discretization is more difficult to implement than random walk simulations, the approach benefits from a long history of theoretical and numerical developments by the mathematical and engineering communities. In particular, software packages for the automated solutions of partial differential equations using finite element discretization, such as FEniCS, are undergoing active support and development. However, because diffusion MRI simulation is a relatively new application area, there is still a gap between the simulation needs of the MRI community and the available tools provided by finite element software packages. In this paper, we address two potential difficulties in using FEniCS for diffusion MRI simulation. First, we simplified software installation by the use of FEniCS containers that are completely portable across multiple platforms. Second, we provide a portable simulation framework based on Python and whose code is open source. This simulation framework can be seamlessly integrated with cloud computing resources such as Google Colaboratory notebooks working on a web browser or with Google Cloud Platform with MPI parallelization. We show examples illustrating the accuracy, the computational times, and parallel computing capabilities. The framework contributes to reproducible science and open-source software in computational diffusion MRI with the hope that it will help to speed up method developments and stimulate research collaborations [11].
- We performed simulations for a collaborative project with Demian Wassermann of the Parietal team on distinguishing between spindle and pyramidal neurons with Multi-shell Diffusion MRI [34].
- We continued in the simulation and modeling of heart diffusion MRI with the post-doc project of Imen Mekkaoui, funded by Inria-EPFL lab. The project is co-supervised with Jan Hesthaven, Chair of Computational Mathematics and Simulation Science (MCSS), EPFL. An article on this topic is under preparation.

5.6. Uncertainty Quantification methods

5.6.1. Acceleration of Domain Decomposition Methods for Stochastic Elliptic Equations

João F. Reis, Olivier P. Le Maître, Pietro M. Congedo, Paul Mycek

We propose a Monte Carlo based method to compute statistics from a solution of a stochastic elliptic equation. Solutions are computed through an iterative solver. We present a parallel construction of a robust stochastic preconditioner to accelerate the iterative scheme. This preconditioner is built before the sampling, at an offline stage, based on a decomposition of the geometric domain. Once constructed, a realisation of the preconditioner is generated for each sample and applied to an iterative method to solve the corresponding deterministic linear system. This approach is not restricted to a single iterative method and can be adapted to different iterative techniques. We demonstrate the efficiency of this approach with extensive numerical results, divided into two examples. The first example is a one-dimensional equation. The reduced dimension of the first example allows the construction of global operators and consequently, an extensive analysis of the convergence and stability properties of the proposed approach. The second example is an analogous two-dimensional version. We demonstrate the performance of the proposed preconditioner by comparison with other deterministic preconditioners based on the median of the coefficient fields. An article on this topic is under preparation.

5.6.2. Clustering based design of experiments for Systems of Gaussian processes

F. Sanson, O.P. Le Maitre, P.M. Congedo

Multi-physics problems in engineering can be often modeled using a System of Solvers (SoS), which is simply a set of solvers coupled together. SoS could be computationally expensive, for example in parametric studies, uncertainty quantification or sensitivity analysis, so typically requiring the construction of a global surrogate model of the SoS to perform such costly analysis. One recurrent strategy in literature consists of building a system of surrogate models where each solver is approximated with a local surrogate model. This approach can be efficient if good training sets for each surrogate can be generated, in particular on the intermediate variables (which are the outputs of an upstream solver and the inputs of a downstream one) that are a priori unknown. In this work, we propose a novel strategy to construct efficient training sets of the intermediate variables, using clustering-based techniques formulated for a systems of Gaussian processes (SoGP). In this way, improved coverage of the intermediate spaces is attained compared to randomly generated training sets. The performances of this approach are assessed on several test-cases showing that the clustering training strategy is systematically more efficient than randomly sampled training points [19].

5.6.3. Extension of AK-MCS for the efficient computation of very small failure probabilities

N. Razaaly, P.M. Congedo

We consider the problem of estimating a probability of failure p_f, defined as the volume of the excursion set of a complex (*e.g.* output of an expensive-to-run finite element model) scalar performance function J below a given threshold, under a probability measure that can be recast as a multivariate standard Gaussian law using an isoprobabilistic transformation. We propose a method able to deal with cases characterized by multiple failure regions, possibly very small failure probability p_f (say $\sim 10^{-6} - 10^{-9}$), and when the number of evaluations of J is limited. The present work is an extension of the popular Kriging-based active learning algorithm known as AK-MCS, permitting to deal with very low failure probabilities. The key idea merely consists in replacing the Monte-Carlo sampling, used in the original formulation to propose candidates and evaluate the failure probability, by a centered isotropic Gaussian sampling in the standard space, which standard deviation is iteratively tuned. This *extreme* AK-MCS (eAK-MCS) inherits its former multi-point enrichment algorithm allowing to add several points at each iteration step, and provide an estimated failure probability based on the Gaussian nature of the Kriging surrogate. Both the efficiency and the accuracy of the proposed method are showcased through its application to two to eight dimensional analytic examples, characterized by very low failure probabilities. Numerical experiments conducted with *unfavorable* initial *Design of Experiment* suggests the ability of the proposed method to detect failure domains [30].

5.6.4. An Efficient Kriging-Based Extreme Quantile Estimation suitable for expensive performance function

N. Razaaly, P.M. Congedo

We propose here a method for fast estimation of the quantiles associated with very small levels of probability, where the scalar performance function J is complex (e.g. output of an expensive-to-run finite element model), under a probability measure that can be recast as a multivariate standard Gaussian law using an isoprobabilistic transformation. A surrogate-based approach (Gaussian Processes) combined with adaptive experimental designs allows to iteratively increase the accuracy of the surrogate while keeping the overall number of J evaluations low. Direct use of Monte-Carlo simulation even on the surrogate model being too expensive, the key idea consists in using an Importance Sampling method based on an isotropic centered Gaussian with large Standard deviation permitting a cheap estimation of the quantiles of the surrogate. Similar to the strategy presented by Schobi and Sudret (2016), the surrogate is adaptively refined using a parallel infill refinement of an algorithm suitable for very small failure probability. We finally elaborate a multi-quantile selection approach allowing to exploit high-performance computing architectures further. We illustrate the performances of the proposed method on several two and six-dimensional cases. Accurate results are obtained with less than 100 evaluations of J. An article on this topic is under preparation.

5.7. Shape optimization under uncertainties

5.7.1. Non-Parametric Measure Approximations for Constrained Multi-Objective Optimisation under Uncertainty

M. Rivier, P.M. Congedo

In this paper, we propose non-parametric estimations of robustness and reliability measures to solve constrained multi-objective optimisation under uncertainty. These approximations with tunable fidelity permit to capture the Pareto front in a parsimonious way, and can be exploited within an adaptive refinement strategy. First, we build a non-Gaussian surrogate model of the objectives and constraints, allowing for more representativeness and detecting potential correlations. Additionally, we illustrate an efficient approach for obtaining joint representations of these robustness and reliability measures, which discriminates more sharply the Pareto-optimal designs from the others. Secondly, we propose an adaptive refinement strategy, using these tunable fidelity approximations to drive the computational effort towards the computation of the optimal area. To this extent, an adapted Pareto dominance rule and Pareto optimal probability computation are formulated. We assess the performance of the proposed strategy on several analytical test-cases against classical approaches in terms of the average distance to the Pareto front. Finally, we illustrate the performances of the method on the case of shape optimisation under uncertainty of an Organic Rankine Cycle turbine [26].

5.7.2. Optimization under Uncertainty of Large Dimensional Problems usingQuantile Bayesian Regression

C. Sabater, O. Le Maitre, P.M. Congedo, S. Goertz

When dealing with robust optimization problems, surrogate models are traditionally constructed to efficiently obtain the statistics of the random variable. However, when a large number of uncertainties are present, the required number of training samples to construct an accurate surrogate generally increases exponentially. The use of surrogates also requires the parametrization of the uncertainties. We present a novel approach for robust design insensitive to the number of uncertainties and that is able to deal with non-parametric uncertainties by leveraging a Bayesian formulation of quantile regression. The method does not require the use of any surrogate in the stochastic space. It is able to globally predict at every design point any given quantile of the random variable. In addition, it provides an estimation of the error in this prediction due to the limited sampling by making use of the posterior distribution of the model parameters.The framework includes an active infill that efficiently balances exploration with exploitation and accelerates the optimization process by increasing the accuracy of the statistic in the regions of interest. We validate the method in test functions and observe good convergence properties towards the determination of the location of the global optima. The framework is applied towards the aerodynamic robust design of an airfoil with a shock control bump under 382 geometrical and operational uncertainties. The framework is able to efficiently find the optimum configuration in complex, large-scale problems. An article on this topic is under preparation.

5.8. Application of Uncertainty Quantification studies to fluid-dynamics problems

5.8.1. Multi-fidelity surrogate-based optimization of transonic and supersonic axial turbines profiles

N. Razaaly, G. Persico, P.M. Congedo

This study presents a multifidelity surrogate-based approach for the optimization of the LS89 high pressure axial turbine vane to significantly reduce the computational cost associated to high fidelity CFD simulations while exploiting models of lower fidelity. A cokriging method is used to simultaneously take into account quantities of interest (QoI) coming from models of different fidelities providing a global surrogate model. A classic bayesian global optimization method permits to iteratively propose desing of interest. It relies on the maximization of the so-called Expected Improvement criterion. A geometrical parametrization technique based on B-splines is considered to describe the profile geometry. The mass-flow rate and the outlet angle are constrained. The optimization study reveals significant reduction in computational cost w.r.t. classic optimization frameworks based on a single fidelity, such as, adjoint-based and gradient-free methods, while providing similar improvements in term of fitness functions [23].

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

- A CIFRE PhD thesis started in April 2017 with Safran Tech. The student is M. Florian Feppon who is working on "topology optimization for a coupled thermal-fluid-structure system”.
- A CIFRE PhD thesis started in October 2017 with Renault. The student is Mrs Lalaina Rakotondrainibe who is working on "topology optimization of connections between mechanical parts”.
- A CIFRE PhD thesis started in January 2019 with Safran Tech. The student is M. Martin Bihr who is working on "Optimisation Topologique du couple support/pièce pour la fabrication additive métallique sur lit de poudre".
- A CIFRE PhD thesis started November 2017 with EDF. The student is H. Girardon who is working on "level set method for eddy current non destructive testting".
- A CIFRE PhD thesis started May 2017 with ArianeGroup. The student is M. Mickael Rivier who is working on "Optimization under uncertainty methods for expensive computer codes".
- A CIFRE PhD thesis started November 2018 with CEA CESTA. The student is M. Paul Novello who is working on "Deep Learning for atmospheric reentry".

6.2. Bilateral Grants with Industry

- The SOFIA project (SOlutions pour la Fabrication Industrielle Additive métallique) started in the summer of 2016. Its purpose is to make research in the field of metallic additive manufacturing. The industrial partners include Michelin, FMAS, ESI, Safran and others. The academic partners are different laboratories of CNRS, including CMAP at Ecole Polytechnique. The project is funded for 6 years by BPI (Banque Publique d’Investissement).
- G. Allaire is participating to the TOP project at IRT SystemX which started in February 2017. It is concerned with the development of a topology optimization platform with industrial partners (Renault, Safran, Airbus, ESI).
- FUI project Saxsize. This three years project started in October 2015 and extended till April 2019 and it involves Xenocs (coordinator), Inria (DEFI), Pyxalis, LNE, Cordouan and CEA. It is a followup of Nanolytix where a focus is put on SAXS quantifications of dense nanoparticle solutions.

- Contract with ArianeGroup, Activity around techniques for Uncertainty Quantification, Coordinator: P.M. Congedo.
- Contract with CEA, Activity around techniques for numerical error estimation and uncertainty quantification, Coordinator: P.M. Congedo.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. EVE

- Title : Virtual prototyping of EVE engines
- Type : Co-funded from Region Aquitaine and Inria
- Duration : 36 months
- Starting : October 2018
- Coordinator : P.M. Congedo
- Abstract : The main objective of this thesis is the construction of a numerical platform, for permitting an efficient virtual prototyping of the EVE expander. This will provide EXOES with a numerical tool, that is much more predictive with respect to the tools currently available and used in EXOES, by respecting an optimal trade-off in terms of complexity/cost needed during an industrial design process.i Two research axes will be mainly developed. First, the objective is to perform some high-predictive numerical simulation for reducing the amount of experiments, thanks to a specific development of RANS tools (Reynolds Averaged Navier-Stokes equations) for the fluids of interest for EXOES. These tools would rely on complex thermodynamic models and a turbulence model that should be modified. The second axis is focused on the integration of the solvers of different fidelity in a multi-fidelity platform for performing optimization under uncertainties. The idea is to evaluate the system performances by using massively the low-fidelity models, and by correcting these estimations via only few calculations with the high-fidelity code.

7.2. European Initiatives

7.2.1. FP7 & H2020 Projects

7.2.1.1. UTOPIAE

Program: H2020 MSCA-ITN

Project acronym: UTOPIAE

Project title: Handling the unknown at the edge of tomorrow

Duration: January 2017- December 2020

Coordinator: M. Vasile (Strathclyde University)

Other partners: see http://utopiae.eu/ for additional details

UTOPIAE is a European research and training network looking at cutting edge methods bridging optimisation and uncertainty quantification applied to aerospace systems. The network will run from 2017 to 2021, and is funded by the European Commission through the Marie Skłodowska-Curie Actions of H2020. The network is made up of 15 partners across 6 European countries, including the UK, and one international partner in the USA, collecting mathematicians, engineers and computer scientists from academia, industry, public and private sectors.

Mission statement : To train, by research and by example, 15 Early Stage Researchers in the field of uncertainty quantification and optimisation to become leading independent researchers and entrepreneurs that will increase the innovation capacity of the EU. To equip the researchers with the skills they will need for successful careers in academia and industry. To develop fundamental mathematical methods and algorithms to bridge the gap between Uncertainty Quantification and Optimisation and between Probability Theory and Imprecise Probability Theory for Uncertainty Quantification to efficiently solve high-dimensional, expensive and complex engineering problems.

7.3. International Initiatives

7.3.1. Inria International Labs

P.M. Congedo is the Inria Coordinator of the CWI-Inria Inria International Lab.

IIL CWI-Inria

Associate Team involved in the International Lab:

7.3.1.1. COMMUNES

Title: Computational Methods for Uncertainties in Fluids and Energy Systems

International Partner (Institution - Laboratory - Researcher):

CWI (Netherlands) - Scientific Computing Group - Daan Crommelin

Start year: 2017

See also: https://project.inria.fr/inriacwi/projects/communes/

This project aims to develop numerical methods capable to take into account efficiently unsteady experimental data, synthetic data coming from numerical simulation and the global amount of uncertainty associated to measurements, and physical-model parameters. We aim to propose novel algorithms combining data-inferred stochastic modeling, uncertainty propagation through computer codes and data assimilation techniques. The applications of interest are both related to the exploitation of renewable energy sources: wind farms and solar Organic Rankine Cycles (ORCs).

7.3.1.2. Informal International Partners

University of Zurich : R. Abgrall. Collaboration on high order adaptive methods for CFD and uncertainty quantification.

Politecnico di Milano, Aerospace Department (Italy) : Pr. A. Guardone. Collaboration on ALE for complex flows (compressible flows with complex equations of state).

von Karman Institute for Fluid Dynamics (Belgium). With Pr. T. Magin we work on Uncertainty Quantification problems for the identification of inflow condition of hypersonic nozzle flows.

Rutgers University. Collaboration with Pr. F. Cakoni on transmission eigenvalues.

University of Delaware. Collaboration with Pr. D. Colton on inverse scattering theory

Ecole Nationale des Ingénieurs de Tunis. Collaboration with Pr. M. Bellasoued on inverse scattering problems

Faculté des Sciences de Sfax. Collaboration with Pr. S. Chaabane on inverse problems for singular parameters

University of Sousse. Collaboration with Pr. M. Khenissi on transmission eigenvalues

Colorado School of Mines. Collaboration with F. Pourahmadian on differential LSM

ns of solution derivatives.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

- Fioralba Cakoni and David Colton, 1 week, March 2019

7.4.1.1. PostDocs, Internships

- PostDoc, Xiaoli Liu, Sampling methods for time dependent problems, H. Haddar
- Master thesis, Marwa Mansouri, Inside-outside duality with artificial backgrounds, L. Chesnel and H. Haddar.
- PostDoc, Imen Mekkaoui, In-vivo cardiac diffusion magnetic resonanace imaging: simulations and parameters estimation, Jing Rebecca Li and Jan Hesthaven.
- Master thesis, Try Nguyen Tran, French-Vietnam Master Program in Applied Mathematics, Jing Rebecca Li
- Master thesis, Nouha jenhani, ENIT, LAMSIN, H. Haddar
- Master thesis, Amal Labidi, ENIT, LAMSIN, H. Haddar

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. General Chair, Scientific Chair

- P.M Congedo is the Chair of the CWI-Inria workshop at CWI in Amsterdam on September 19, 20 2019.
- P.M. Congedo is the Chair of UQOP2020 Conference, organized in Paris on March 18-21, 2019.

8.1.2. Member of the Organizing Committees

- L. Chesnel co-organizes the Journée de rentrée (2019) of the Centre de Mathématiques Appliquées of École Polytechnique.
- L. Chesnel co-organizes the seminar of the Centre de Mathématiques Appliquées of École Polytechnique.
- L. Chesnel co-organizes the seminar of the Inria teams Defi-M3DISIM-Poems.
- M. Bonazzoli organizes the working group of Defi team.
- J.R. Li is organizer of Ecole d'ete d'excellence for Chinese Master's students funded by French Embassy in China, 07/2019.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

- G. Allaire is member of the editorial boards of
  - Book series "Mathématiques et Applications" of SMAI and Springer,
  - ESAIM/COCV, Structural and Multidisciplinary Optimization,
  - Discrete and Continuous Dynamical Systems Series B,
  - Computational and Applied Mathematics,
  - Mathematical Models and Methods in Applied Sciences (M3AS),
  - Annali dell'Universita di Ferrara,
  - OGST (Oil and Gas Science and Technology),
  - Journal de l'Ecole Polytechnique - Mathématiques,
  - Journal of Optimization Theory and Applications.
- P.M. Congedo is Editor of Mathematics and Computers in Simulation, MATCOM (Elsevier).
- H. Haddar is member of the editorial boards of
  - Inverse Problems
  - SIAM Journal on Scientific Computing
  - SIAM Journal of Mathematical Analysis

8.1.3.2. Reviewer - Reviewing Activities

We reviewed papers for top international journals in the main scientific themes of the team.

8.1.4. Invited Talks

- G. Allaire
  - SIAM Geosciences, Houston, March 11-14, 2019.
  - DCAMM seminar, DTU, Copenhagen, April 5, 2019.
  - WCSMO, Beijing, May 20-24, 2019.
  - Mathematical Design of New Materials, Cambridge, June 3-14, 2019.
  - Chalmers Colloquium, Sweden, August 26-30, 2019.
  - Sim-AM, Pavia, September 11-13, 2019.
  - Shape Optimization and Isoperimetric and Functional Inequalities, Levico Terme, September 23-27, 2019.
  - Computational modelling of Complex Materials across the Scales, Glasgow, October 1-4, 2019.
  - New trends in PDE constrained optimization, Linz, October 15-18, 2019.
- M. Bonazzoli
  - Seminar at POEMS lab, ENSTA-ParisTech, Palaiseau, France.
  - ENUMATH 2019, European Numerical Mathematics and Advanced Applications Conference, Egmond aan Zee, Netherlands.
  - Parallel Solution Methods for Systems Arising from PDEs, Marseille, France (Plenary invited talk).
- L. Chesnel
  - Waves conference, Vienna, August 2019.
  - Applied Inverse Problems conference, Grenoble, 2019.
- H. Haddar
  - Applied Inverse Problems conference, Grenoble, July 2019
  - International Conference on Antenna Measurements & Applications, Bali, October 2019
  - New Trends in Analysis and Probability, Sousse, September 2019
  - Workshop in the memory of A. Lechleiter, Bremen, May 2019
  - La journée des rencontres DEFI-MEDISIM-POEMS, December, 2019
- P.M. Congedo
  - Workshop "Numerical simulation of hypersonic flows, July 8, 2019.
  - Seminar at ONERA, Meudon, November 29, 2019.

8.1.5. Leadership within the Scientific Community

- G. Allaire is a board member of Institut Henri Poincaré (IHP). He is the chairman of the scientific council of IFPEN (French Petroleum Institute and New Energies). He is the chairman of the scientific council of AMIES (Agency for Interaction in Mathematics with Business and Society).

8.1.6. Scientific Expertise

- G. Allaire is a member of the "comité national" CNRS, section 41 (mathematics).
- G. Allaire is a member of the scientific board of the Gaspard Monge program on optimization (PGMO) at the Jacques Hadamard Mathematical Foundation.
- J.R. Li is Member of the SIAM Committee on Programs and Conferences 2017-2019.
- J.R. Li is Member Elu of Inria Commission d'Evaluation, 2015-2019.
- M. Bonazzoli was a member of the Evaluation committee for the 2020 call of Inria Associate Teams programme.

- H. Haddar was the president of the evaluation committee for mathemùatical laboratries at the Universities of Sfax and Sousse (Tunisia)

8.1.7. Research Administration

- J.R. Li is correspondant International for Centre de Mathematiques Appliquees, Ecole Polytechnique, 2018-present.
- J.R. Li is responsable for the Ecole Polytechnique part of the French-Vietnam Master Program in Applied Mathematics, 2016-present.
- M. Bonazzoli is the International partnerships Scientific Correspondent for Inria Saclay.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

- Master: Grégoire Allaire, Approximation Numérique et Optimisation, for students in the second year of Ecole Polytechnique curriculum: 8 lessons of 1h30.
- Master: Grégoire Allaire, Transport and diffusion, for students in the third year of Ecole Polytechnique curriculum. 9 lessons of 2h jointly with F. Golse.
- Master: Houssem Haddar, Waves and imaging: Concepts, Theory and Applications, Master M2 ”mathematical modeling”: 9 lessons of 3h.
- Master: Houssem Haddar, Inverse scattering problems, Master M2, ENIT, 10 lessons of 3h.
- Master: Lucas Chesnel, Elementary tools of analysis for partial differential equations, for students in the first year of Ensta ParisTech curriculum, 25 equivalent TD hours.
- Master: Lucas Chesnel, Numerical approximation and optimisation, for students in the second year of Ecole Polytechnique curriculum: 2 TDs of 4h + one project.
- Master: Lucas Chesnel, Modal Modélisation mathématique par la démarche expérimentale, for students in the second year of Ecole Polytechnique curriculum: 5 TDs of 2h.
- Master: Grégoire Allaire, Optimal design of structures, for students in the third year of Ecole Polytechnique curriculum. 9 lessons of 1h30.
- Master: Grégoire Allaire, Theoretical and numerical analysis of hyperbolic systems of conservation laws, Master M2 ”mathematical modeling”, 8 lessons of 3h.
- Master: Jing Rebecca Li, Lecturer of course Mathematical and numerical foundations of modeling and simulation using partial differential equations French-Vietnam Master in Applied Mathematics, University of Science, Ho Chi Minh City, 9/2019. 2 weeks.
- Master: P.M. Congedo, Numerical methods in Fluid Mechanics, ENSTA ParisTech, 12 h.
- Master: P.M. Congedo, Numerical methods for Hyperbolic Problems, von Karman Institute for Fluid Dynamics, 12 h.
- Doctorat: Houssem Haddar, Inverse problems, Executive Education, Ecole Polytechnique, 9h

8.2.2. Supervision

- PhD: K. Napal, On the use of sampling methods and spectral signatures for the resolution of inverse scattering problems (defended, December 2019), L. Audibert, L. Chesnel and H. Haddar.
- PhD: F. Feppon (defended, December 2019) sur l’optimisation topologique de systèmes couplés fluide-solide-thermique, G. Allaire and Ch. Dapogny.
- PhD: M. Kchaw, Higher order homogenization tensors for DMRI modeling (defended July 2019), H. Haddar and M. Moakher
- PhD: B. Charfi, Identification of the singular support of generalized impedance boundary conditions (defended September 2019), S. Chaabane and H. Haddar.

- PhD: F. Sanson, Estimation du risque humain lié à la retombée d'objets spatiaux sur Terre (defended in September 2019), P.M. Congedo, O. Le Maitre.
- PhD: N. Razaaly, Rare Event Estimation and Robust Optimization Methods with Applications to ORC Turbine Cascade (defended in July 2019), P.M. Congedo.
- PhD: G. Gori, Non-ideal compressible-fluid dynamics: developing a combined perspective on modeling, numerics and experiments (defended in January 2019), A. Guardone, P.M. Congedo.
- PhD: J. Carlier, Schémas aux résidus distribués et méthodes à propagation des ondes pour la simulation d'écoulements compressibles diphasiques avec transfert de chaleur et de masse (defended in December 2019), M. Pelanti, P.M. Congedo.
- Ph.D. in progress: S. Houbar sur la cavitation dans le fluide caloporteur induite par les mouvements des assemblages d'un réacteur (CEA, to be defended in 2020). G. Allaire and G. Campioni.
- Ph.D. in progress: M. Boissier, Optimisation couplée de la topologie des formes et de la trajectoire de lasage en fabrication additive (to be defended in 2020). G. Allaire and Ch. Tournier.
- L. Rakotondrainibe sur l'optimisation des liaisons enre pièces dans les système mécaniques (Renault, to be defended in 2020). G. Allaire.
- J. Desai sur l'optimisation topologique de structures au comportement non-linéaire avec des méthodes de déformation de maillage (IRT SystemX, to be defended in 2021). G. Allaire and F. Jouve,
- PhD in progress: H. Girardon, Non destructive testing of PWR tubes using eddy current rotating coils, to be defended in 2021, H. Haddar and L. Audibert
- PhD in progress: M. Rihani, Maxwell's equations in presence of metamaterials (to be defended in 2021), A.-S. Bonnet-BenDhia and L. Chesnel.
- PhD in progress: Chengran Fang, Enabling cortical cell-specific sensitivity on clinical multi-shell diffusion MRI microstructure measurements. (to be defended in 2022) Jing Rebecca Li and Demian Wassermann
- PhD in progress: Nouha Jenhani, Differential sampling methods for defect imaging in periodic layers. (to be defended in 2022) Houssem Haddar and Mourad Bellasoued
- PhD in progress: Amal Labidi, Inverse problems for wave equation with magnetic potential. (to be defended in 2022) Houssem Haddar and Mourad Bellasoued
- PhD in progress: Marwa Mansouri, Inside outside duality for artificial backgrounds. (to be defended in 2022) Houssem Haddar, Lucas Chesnel and Moez Khenissi
- PhD in progress: M. Bihr sur la fabrication additive et l'optimisation topologique de structures (to be defended in 2022) G. Allaire and B. Bogosel.
- PhD in progress: R. Delvaux sur les algorithmes de couplage à convergence super-linéaire entre neutronique, thermohydraulique et thermique (to be defended in 2022) G. Allaire and C. Patricot,.
- PhD in progress: A. Touiti sur l'optimisation de l'anisotropie pour des structures issues de la fabrication additive (to be defended in 2022) G. Allaire and F. Jouve.
- PhD in progress: M. Rivier, optimization under uncertainty through a Bounding-Box concept (to be defended in May 2020), P.M. Congedo.
- PhD in progress: Joao Reis, Advanced methods for stochastic elliptic PDEs (to be defended in October 2020), P.M. Congedo, O. Le Maitre.
- PhD in progress: Anabel Del Val, Advanced bayesian methods for aerospace applications (to be defended in October 2020), P.M. Congedo, O. Le Maitre, O. Chazot, T. Magin.
- PhD in progress: P. Novello, Deep learning for reentry atmosperic flows (to be defended in November 2021), P.M. Congedo, D. Lugato, G. Poette.
- PhD in progress: E. Solai, Virtual Prototyping of the EVE expander (to be defended in October 2021), P.M. Congedo, H. Beaugendre.

- PhD in progress: N. Leoni, Bayesian inference of model error in imprecise models (to be defended in February 2022), P.M. Congedo, O. Le Maitre, M.G. Rodio.

8.3. Popularization

- M. Bonazzoli and L. Chesnel made a presentation in the context of the Fête de la science 2019 to several groups of young students (from 10 to 17 years old).
- M. Bonazzoli was representative for SMAI at the Métiers des Maths stand at the 20th Salon Culture et Jeux Mathématiques.

8.3.1. Internal or external Inria responsibilities

- P.M. Congedo is Deputy Coordinator of "Maths/Engineering" Program of the Labex Mathématiques Hadamard.
- J.R. Li is Member Elu of Inria Commission d'Evaluation, 2015-2019.
- M. Bonazzoli is the International partnerships Scientific Correspondent for Inria Saclay.

8.3.2. Internal action

- M. Bonazzoli, H. Haddar and J.R. Li monitored an internship at Defi team for 6 middle school students (one afternoon).

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] L. AUDIBERT, L. CHESNEL, H. HADDAR. *Inside-outside duality with artificial backgrounds*, in "Inverse Problems", September 2019, vol. 35, n^o 10, 104008, https://hal.archives-ouvertes.fr/hal-02095660

[2] M. BAKRY, O. BUNĂU, H. HADDAR. *A robust expectation-maximization method for the interpretation of small-angle scattering data from dense nanoparticle samples*, in "Journal of Applied Crystallography", October 2019, vol. 52, n^o 5, p. 926-936 [*DOI :* 10.1107/S1600576719009373], https://hal.inria.fr/hal-02416529

[3] M. BONAZZOLI, V. DOLEAN, I. G. GRAHAM, E. A. SPENCE, P.-H. TOURNIER. *Domain decomposition preconditioning for the high-frequency time-harmonic Maxwell equations with absorption*, in "Mathematics of Computation", May 2019, https://arxiv.org/abs/1711.03789 , forthcoming [*DOI :* 10.1090/MCOM/3447], https://hal.archives-ouvertes.fr/hal-01644011

[4] L. BOURGEOIS, L. CHESNEL. *On quasi-reversibility solutions to the Cauchy problem for the Laplace equation: regularity and error estimates*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2019, forthcoming [*DOI :* 10.1051/M2AN/2019073], https://hal.inria.fr/hal-02385487

[5] L. BOURGEOIS, L. CHESNEL, S. FLISS. *On well-posedness of time-harmonic problems in an unbounded strip for a thin plate model*, in "Communications in Mathematical Sciences", December 2019, https://hal.archives-ouvertes.fr/hal-01883560

[6] L. CHESNEL, S. NAZAROV, J. TASKINEN. *Surface waves in a channel with thin tunnels and wells at the bottom: non-reflecting underwater topography*, in "Asymptotic Analysis", September 2019, https://hal.archives-ouvertes.fr/hal-02426417

[7] L. CHESNEL, V. PAGNEUX.*From zero transmission to trapped modes in waveguides*, in "Journal of Physics A: Mathematical and Theoretical", March 2019, https://hal.archives-ouvertes.fr/hal-01952980

[8] G. GORI, M. ZOCCA, G. CAMMI, A. SPINELLI, P. M. CONGEDO, A. GUARDONE.*Accuracy assessment of the Non-Ideal Computational Fluid Dynamics model for siloxane MDM from the open-source SU2 suite*, in "European Journal of Mechanics - B/Fluids", 2019, vol. 79, p. 109-120 [*DOI :* 10.1016/J.EUROMECHFLU.2019.08.014], https://hal.archives-ouvertes.fr/hal-02389668

[9] H. HADDAR, A. KONSCHIN.*Factorization Method for Imaging a Local Perturbation in Inhomogeneous Periodic Layers from Far Field Measurements*, in "Inverse Problems and Imaging ", 2019, forthcoming, https://hal.inria.fr/hal-02416496

[10] J.-R. LI, V.-D. NGUYEN, T. N. TRAN, J. VALDMAN, C.-B. TRANG, K. V. NGUYEN, D. T. S. VU, H. A. TRAN, H. T. A. TRAN, T. M. P. NGUYEN.*SpinDoctor: A MATLAB toolbox for diffusion MRI simulation*, in "NeuroImage", November 2019, vol. 202, 116120 [*DOI :* 10.1016/J.NEUROIMAGE.2019.116120], https://hal.archives-ouvertes.fr/hal-02431597

[11] V.-D. NGUYEN, M. DE LEONI, T. DANCHEVA, J. JANSSON, J. HOFFMAN, D. WASSERMANN, J.-R. LI.*Portable simulation framework for diffusion MRI*, in "Journal of Magnetic Resonance", December 2019, vol. 309, 106611 [*DOI :* 10.1016/J.JMR.2019.106611], https://hal.archives-ouvertes.fr/hal-02431598

[12] V.-D. NGUYEN, J. JANSSON, H. T. A. TRAN, J. HOFFMAN, J.-R. LI.*Diffusion MRI simulation in thin-layer and thin-tube media using a discretization on manifolds*, in "Journal of Magnetic Resonance", February 2019, vol. 299, p. 176-187 [*DOI :* 10.1016/J.JMR.2019.01.002], https://hal.archives-ouvertes.fr/hal-02431595

[13] K. V. NGUYEN, D. LE BIHAN, L. CIOBANU, J.-R. LI.*The time-dependent diffusivity in the abdominal ganglion of Aplysia californica: experiments and simulations*, in "Biomedical Physics & Engineering Express", July 2019, vol. 5, n^{o} 4, 045036 [*DOI :* 10.1088/2057-1976/AB301E], https://hal.archives-ouvertes.fr/hal-02431596

[14] N. RAZAALY, D. CROMMELIN, P. M. CONGEDO.*Efficient Estimation of Extreme Quantiles using Adaptive Kriging and Importance Sampling*, in "International Journal for Numerical Methods in Engineering", December 2019 [*DOI :* 10.1002/NME.6300], https://hal.inria.fr/hal-02436230

[15] M. RIVIER, P. M. CONGEDO.*Surrogate-assisted Bounding-Box approach for optimization problems with tunable objectives fidelity*, in "Journal of Global Optimization", January 2019 [*DOI :* 10.1007/S10898-019-00823-9], https://hal.inria.fr/hal-02436229

[16] M. RIVIER, J. LACHAUD, P. M. CONGEDO.*Ablative thermal protection system under uncertainties including pyrolysis gas composition*, in "Aerospace Science and Technology", January 2019 [*DOI :* 10.1016/J.AST.2018.11.048], https://hal.inria.fr/hal-02436221

[17] F. SANSON, J.-M. BOUILLY, C. BERTORELLO, P. M. CONGEDO.*Breakup prediction under uncertainty: Application to upper stage controlled reentries from GTO orbit*, in "Aerospace Science and Technology", January 2019 [*DOI :* 10.1016/J.AST.2019.02.031], https://hal.inria.fr/hal-02436222

[18] F. SANSON, C. FRUEH.*Quantifying uncertainties in signal position in non-resolved object images: application to space object observation*, in "Advances in Space Research", 2019, forthcoming [*DOI :* 10.1016/J.ASR.2018.12.040], https://hal.archives-ouvertes.fr/hal-01968808

[19] F. SANSON, O. LE MAITRE, P. M. CONGEDO.*Systems of Gaussian process models for directed chains of solvers*, in "Computer Methods in Applied Mechanics and Engineering", August 2019, vol. 352, p. 32-55 [*DOI :* 10.1016/J.CMA.2019.04.013], https://hal.inria.fr/hal-02327697

[20] P.-H. TOURNIER, I. ALIFERIS, M. BONAZZOLI, M. DE BUHAN, M. DARBAS, V. DOLEAN, F. HECHT, P. JOLIVET, I. EL KANFOUD, C. MIGLIACCIO, F. NATAF, C. PICHOT, S. SEMENOV.*Microwave Tomographic Imaging of Cerebrovascular Accidents by Using High-Performance Computing*, in "Parallel Computing", July 2019, vol. 85, p. 88-97, https://arxiv.org/abs/1607.02573 [*DOI :* 10.1016/J.PARCO.2019.02.004], https://hal.archives-ouvertes.fr/hal-01343687

[21] T. YU, J.-F. CHAIX, L. AUDIBERT, D. KOMATITSCH, V. GARNIER, J.-M. HENAULT.*Simulations of ultrasonic wave propagation in concrete based on a two-dimensional numerical model validated analytically and experimentally*, in "Ultrasonics", February 2019, vol. 92, p. 21 - 34 [*DOI :* 10.1016/J.ULTRAS.2018.07.018], https://hal.archives-ouvertes.fr/hal-01934352

International Conferences with Proceedings

[22] N. RAZAALY, G. GORI, G. IACCARINO, P. M. CONGEDO.*Optimization of an ORC supersonic nozzle under epistemic uncertainties due to turbulence models*, in "GPPS 2019 - Global Power and Propulsion Society", Zurich, Switzerland, January 2019, https://hal.inria.fr/hal-01982227

[23] N. RAZAALY, G. PERSICO, P. M. CONGEDO.*Optimal Design of ORC Turbine Blades Under Geometric and Operational Uncertainties*, in "ORC 2019 - International Seminar on ORC Power Systems 2019", Athens, Greece, September 2019, https://hal.inria.fr/hal-02389654

[24] M. RIVIER, P. M. CONGEDO.*A Surrogate-Assisted Multi-fidelity Measure Approximation Framework for Efficient Constrained Multiobjective Optimization Under Uncertainty*, in "UNCECOMP 2019 - 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering", Hersonissos, Greece, June 2019, https://hal.inria.fr/hal-02286156

[25] M. RIVIER, P. M. CONGEDO.*Different Measure Approximations for Efficient Constrained Multi-Objective Optimization under Uncertainty*, in "UQOP - Uncertainty Quantification and OPtimization", Paris, France, March 2019, https://hal.inria.fr/hal-02286145

National Conferences with Proceeding

[26] M. RIVIER, P. M. CONGEDO.*Different Measure Approximations for Efficient Constrained Multi-Objective Optimization under Uncertainty*, in "SMAI 2019 - 9e Biennale Française des Mathématiques Appliquées et Industrielles", Guidel, France, May 2019, https://hal.inria.fr/hal-02286141

Conferences without Proceedings

[27] F. SANSON, C. BERTORELLO, J.-M. BOUILLY, P. M. CONGEDO.*Space debris reentry prediction and ground risk estimation using a probabilistic breakup model*, in "AIAA Scitech 2019 Forum", San Diego, United States, American Institute of Aeronautics and Astronautics, January 2019 [*DOI :* 10.2514/6.2019-2234], https://hal.archives-ouvertes.fr/hal-02392632

[28] F. SANSON, O. LE MAITRE, P. M. CONGEDO.*Uncertainty propagation in multiphysics systems of solvers: application to robust space object reentry predictions*, in "UQOP - Conference on Uncertainty Quantification and Optimization", Paris, France, March 2019, https://hal.archives-ouvertes.fr/hal-02392640

Scientific Books (or Scientific Book chapters)

[29] L. AUDIBERT, L. CHESNEL, H. HADDAR, K. NAPAL.*Detecting Sound Hard Cracks in Isotropic Inhomogeneities*, in "ICAV 2018 - Proceedings of the Second International Conference on Acoustics and Vibration", Applied Condition Monitoring, Springer, September 2019, p. 61-73, https://hal.archives-ouvertes.fr/hal-01945670

Research Reports

[30] N. RAZAALY, P. M. CONGEDO.*Extension of AK-MCS for the efficient computation of very small failure probabilities*, Inria Saclay - Equipe DeFI, August 2019, n° RR-9286, https://hal.inria.fr/hal-02265311

Other Publications

[31] M. AUSSAL, Y. BOUKARI, H. HADDAR.*Data Completion Method For the Helmholtz Equation Via Surface Potentials for Partial Cauchy Data*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02416548

[32] R. BUNOIU, L. CHESNEL, K. RAMDANI, M. RIHANI.*Homogenization of Maxwell's equations and related scalar problems with sign-changing coefficients*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02421312

[33] L. CHESNEL, S. NAZAROV.*Exact zero transmission during the Fano resonance phenomenon in non symmetric waveguides*, July 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02189311

[34] V. MENON, G. GUILLERMO, M. PINSK, V.-D. NGUYEN, J.-R. LI, W. CAI, D. WASSERMANN.*Quantitative modeling links in vivo microstructural and macrofunctional organization of human and macaque insular cortex, and predicts cognitive control abilities*, January 2020, working paper or preprint [*DOI :* 10.1101/662601], https://hal.inria.fr/hal-02434382

[35] F. POURAHMADIAN, H. HADDAR.*Differential tomography of micromechanical evolution in elastic materials of unknown micro/macrostructure*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02416504

[36] M. RIVIER, P. M. CONGEDO.*Different measure approximations for efficient constrained multi-objective optimization under uncertainty*, February 2019, SIAM CSE19 - Conference on Computational Science and Engineering, Poster, https://hal.inria.fr/hal-02286159

Project-Team DISCO

Dynamical Interconnected Systems in COmplex Environments

IN COLLABORATION WITH: Laboratoire des signaux et systèmes (L2S)

IN PARTNERSHIP WITH:
CNRS

CentraleSupélec

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization and control of dynamic systems

Table of contents

Project-Team DISCO

Creation of the Team: 2010 January 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization

Other Research Topics and Application Domains:

B2.2.3. - Cancer
B2.3. - Epidemiology
B2.5. - Handicap and personal assistances
B3.6. - Ecology
B4.3.1. - Biofuels
B5.2.3. - Aviation
B7.2.1. - Smart vehicles

1. Team, Visitors, External Collaborators

Research Scientists

Catherine Bonnet [Team leader, Inria, Senior Researcher, HDR]
Islam Boussaada [IPSA, CNRS, Researcher, from Sep 2019, HDR]
Frédéric Mazenc [Inria, Senior Researcher, HDR]

Faculty Members

Silviu-Iulian Niculescu [CNRS, Professor, HDR]
Giorgio Valmorbida [Centrale-Supélec, Associate Professor]

External Collaborator

Ali Zemouche [Univ de Lorraine]

PhD Students

Amina Benarab [IPSA, PhD Student, from Oct 2019]
Caetano Cardeliquio [UNICAMP & Inria, PhD Student, until Oct 2019]
Ali Diab [CNRS & Centralesupelec, PhD Student, from Oct 2019]
Leonardo Broering Groff [CNRS & Centralesupelec, PhD Student]
Jose Hernandez Diez [CNRS & San Luis Potosi, PhD Student]
Alejandro Martinez Gonzalez [CNRS & San Luis Potosi, PhD Student]

Post-Doctoral Fellows

Da Jung Cho [Inria, Post-Doctoral Fellow, from May 2019 until Aug 2019]
Guilherme Mazanti [IPSA, Post-Doctoral Fellow, from Oct 2019]

Visiting Scientist

Yutaka Yamamoto [Kyoto University,30 Oct-8 Dec 2020]

Administrative Assistants

Katia Evrat [Inria, Administrative Assistant, until Oct 2019]

Irina Lahaye [Inria, Administrative Assistant, from Nov 2019]

2. Overall Objectives

2.1. Objectives

The goal of the project is to better understand and well formalize the effects of complex environments on the dynamics of the interconnections, as well as to develop new methods and techniques for the analysis and control of such systems.

It is well-known that the interconnection of dynamic systems has as consequence an increased complexity of the behavior of the total system.

In a simplified way, as the concept of dynamics is well-understood, the interconnections can be seen as associations (by connections of materials or information flows) of distinct systems to ensure a pooling of the resources with the aim of obtaining a better operation with the constraint of continuity of the service in the event of a fault. In this context, the environment can be seen as a collection of elements, structures or systems, natural or artificial constituting the neighborhood of a given system. The development of interactive games through communication networks, control from distance (e.g. remote surgical operations) or in hostile environment (e.g. robots, drones), as well as the current trend of large scale integration of distribution (and/or transport and/or decision) and open information systems with systems of production, lead to new modeling schemes in problems where the dynamics of the environment have to be taken into account.

In order to tackle the control problems arising in the above examples, the team investigates new theoretical methods, develops new algorithms and implementations dedicated to these techniques.

3. Research Program

3.1. Analysis of interconnected systems

The major questions considered are those of the characterization of the stability (also including the problems of sensitivity compared to the variations of the parameters) and the determination of stabilizing controllers of interconnected dynamic systems. In many situations, the dynamics of the interconnections can be naturally modelled by systems with delays (constant, distributed or time-varying delays) possibly of fractional order. In other cases, partial differential equations (PDE) models can be better represented or approximated by using systems with delays. Our expertise on this subject, on both time and frequency domain methods, allows us to challenge difficult problems (e.g. systems with an infinite number of unstable poles).

- Robust stability of linear systems

  Within an interconnection context, lots of phenomena are modelled directly or after an approximation by delay systems. These systems may have constant delays, time-varying delays, distributed delays ...

  For various infinite-dimensional systems, particularly delay and fractional systems, input-output and time-domain methods are jointly developed in the team to characterize stability. This research is developed at four levels: analytic approaches (H_∞-stability, BIBO-stablity, robust stability, robustness metrics) [1], [2], [6], [7], symbolic computation approaches (SOS methods are used for determining easy-to-check conditions which guarantee that the poles of a given linear system are not in the closed right half-plane, certified CAD techniques), numerical approaches (root-loci, continuation methods) and by means of softwares developed in the team [6], [7].

- Robustness/fragility of biological systems

Deterministic biological models describing, for instance, species interactions, are frequently composed of equations with important disturbances and poorly known parameters. To evaluate the impact of the uncertainties, we use the techniques of designing of global strict Lyapunov functions or functional developed in the team.

However, for other biological systems, the notion of robustness may be different and this question is still in its infancy (see, e.g. [57]). Unlike engineering problems where a major issue is to maintain stability in the presence of disturbances, a main issue here is to maintain the system response in the presence of disturbances. For instance, a biological network is required to keep its functioning in case of a failure of one of the nodes in the network. The team, which has a strong expertise in robustness for engineering problems, aims at contributing at the develpment of new robustness metrics in this biological context.

3.2. Stabilization of interconnected systems

- Linear systems: Analytic and algebraic approaches are considered for infinite-dimensional linear systems studied within the input-output framework.

  In the recent years, the Youla-Kučera parametrization (which gives the set of all stabilizing controllers of a system in terms of its coprime factorizations) has been the cornerstone of the success of the H_∞-control since this parametrization allows one to rewrite the problem of finding the optimal stabilizing controllers for a certain norm such as H_∞ or H_2 as affine, and thus, convex problem.

  A central issue studied in the team is the computation of such factorizations for a given infinite-dimensional linear system as well as establishing the links between stabilizability of a system for a certain norm and the existence of coprime factorizations for this system. These questions are fundamental for robust stabilization problems [1], [2].

  We also consider simultaneous stabilization since it plays an important role in the study of reliable stabilization, i.e. in the design of controllers which stabilize a finite family of plants describing a system during normal operating conditions and various failed modes (e.g. loss of sensors or actuators, changes in operating points). Moreover, we investigate strongly stabilizable systems, namely systems which can be stabilized by stable controllers, since they have a good ability to track reference inputs and, in practice, engineers are reluctant to use unstable controllers especially when the system is stable.

- Nonlinear systems

  In any physical systems a feedback control law has to account for limitation stemming from safety, physical or technological constraints. Therefore, any realistic control system analysis and design has to account for these limitations appearing mainly from sensors and actuators nonlinearities and from the regions of safe operation in the state space. This motivates the study of linear systems with more realistic, thus complex, models of actuators. These constraints appear as nonlinearities as saturation and quantization in the inputs of the system [10].

  The project aims at developing robust stabilization theory and methods for important classes of nonlinear systems that ensure good controller performance under uncertainty and time delays. The main techniques include techniques called backstepping and forwarding, contructions of strict Lyapunov functions through so-called "strictification" approaches [4] and construction of Lyapunov-Krasovskii functionals [5], [6], [7] or or Lyapunov functionals for PDE systems [9].

3.3. Synthesis of reduced complexity controllers

- PID controllers

  Even though the synthesis of control laws of a given complexity is not a new problem, it is still open, even for finite-dimensional linear systems. Our purpose is to search for good families of “simple” (e.g. low order) controllers for infinite-dimensional dynamical systems. Within our approach, PID candidates are first considered in the team [2], [60].

For interconnected systems appearing in teleoperation applications, such as the steer-by-wire, Proportional-Derivative laws are simple control strategies allowing to reproduce the efforts in both ends of the teleoperation system. However, due to delays introduced in the communication channels these strategies may result in loss of closed loop stability or in performance degradation when compared to the system with a mechanical link (no communication channel). In this context we search for non-linear proportional and derivative gains to improve performance. This is assessed in terms of reduction of overshoot and guaranteed convergence rates.

- Delayed feedback

  Control systems often operate in the presence of delays, primarily due to the time it takes to acquire the information needed for decision-making, to create control decisions and to execute these decisions. Commonly, such a time delay induces desynchronizing and/or destabilizing effects on the dynamics. However, some recent studies have emphasized that the delay may have a stabilizing effect in the control design. In particular, the closed-loop stability may be guaranteed precisely by the existence of the delay. The interest of considering such control laws lies in the simplicity of the controller as well as in its easy practical implementation. It is intended by the team members to provide a unified approach for the design of such stabilizing control laws for finite and infinite dimensional plants [3], [8].

- Finite Time and Interval Observers for nonlinear systems

  We aim to develop techniques of construction of output feedbacks relying on the design of observers. The objectives pertain to the design of robust control laws which converge in finite time, the construction of intervals observers which ensure that the solutions belong to guaranteed intervals, continuous/discrete observers for systems with discrete measurements and observers for systems with switches.

Finally, the development of algorithms based on both symbolic computation and numerical methods, and their implementations in dedicated Scilab/Matlab/Maple toolboxes are important issues in the project.

4. Application Domains

4.1. Analysis and Control of life sciences systems

The team is involved in life sciences applications. The two main lines are the analysis of bioreactors models (microorganisms; bacteria, microalgae, yeast, etc..) and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML) in collaboration with St Antoine Hospital in Paris. A recent subject is the modelling of epidemics for tropical diseases.

4.2. Energy Management

The team is interested in Energy management and considers control problems in energy networks.

5. New Results

5.1. Multiplicity-induced-dominancy

Participants: Islam Boussaada, Guilherme Mazanti, Hugues Mounier [L2S, CentraleSupélec], Silviu-Iulian Niculescu.

The effects of multiplicity of spectral values on the exponential stability of reduced-order retarded differential equation were emphasized in recent works. In [13] the general class of second-order retarded differential equations is studied. A parametric multiplicity-induced-dominancy (MID) property is characterized, allowing a delayed stabilizing design with reduced complexity. The proposed approach is merely a delayed-output-feedback where the candidates' delays and gains result from the manifold defining the maximal multiplicity of a real spectral value, then, the dominancy is shown using the argument principle.

The work [52] considers retarded differential equations of arbitrary order with a single delay. The existence of a real root with maximal multiplicity is characterized in terms of the equation parameters. This root is shown to be always strictly dominant, determining thus the asymptotic behavior of the system. The dominancy proof is based on improved a priori bounds on the imaginary part of roots on the complex right half-plane and a suitable factorization of the characteristic function, which is an alternative technique to the argument principle.

In [53] we extend the MID property to a given pair of complex conjugate roots for a generic second-order retarded differential equation. Necessary and sufficient conditions for the existence of such a pair are provided, and it is also shown that such a pair is always necessarily dominant. It appears also that when the frequency corresponding to this pair of roots tends to 0, then the pair of roots collapse into a real root of maximal multiplicity. The latter property is exploited in the dominancy proof together with a study of crossing imaginary roots.

In [41] a control-oriented model of torsional vibrations occurring in rotary drilling process is proposed. More precisely, a wave equations with weak damping term is considered. An appropriate stabilizing controller with a reduced number of parameters is proposed for damping such torsional vibrations. Such a controller allows to further explore the effect of multiple roots with maximal admissible multiplicity for linear neutral system with a single delay.

5.2. Pole placement techniques for time-delay systems

Participants: Islam Boussaada, Silviu-Iulian Niculescu, Sami Tliba [L2S, CentraleSupélec], Fazia Bedouhene [University Mouloud Mammeri de Tizi Ouzou, Algeria].

The interest in investigating multiple roots does not rely on the multiplicity itself, but rather on its connection with the dominance of this root, and the corresponding applications in stability analysis and control design. The work [31] extends the analytical characterization of the spectral abscissa for retarded time-delay system with real spectral values which are not necessarily multiple. The effect of the coexistence of such non oscillatory modes on the asymptotic stability of the trivial solution is explored. In particular, the coexistence of an appropriate number of real spectral values makes them rightmost-roots of the corresponding quasipolynomial. Furthermore, if they are negative, this guarantees the asymptotic stability of the trivial solution. As an application of the proposed rightmost-characteristic root assignment, the problem of the vibration damping for a thin flexible beam is considered in [33]. The considered beam is equipped with two piezoelectric patches: one of them works as a sensor and the other as an actuator. Each one of them is bonded on one side of the beam and both are collocated. The model of this system is obtained numerically thanks to a finite element modeling, leading to a linear state-space model. The design of a controller based on proportional and delayed-proportional actions on the input and output signals is proposed, where the properties of the proposed output feedback controller in terms of vibration damping and of robustness is discussed.

5.3. PID tuning for controlling delayed dynamics

Participants: Jie Chen [City University of Hong Kong], Jianqi Chen [City University of Hong Kong], Andong Liu [City University of Hong Kong], Dan Ma [Northeastern University, China], Silviu Niculescu.

Nowadays, the PID controller is the most used in controlling industrial processes. In [21] such a controller is further explored in the control of delayed dynamics. An explicit lower bound on the delay margin of second-order unstable delay systems achievable by PID control is established, which provides a priori a guaranteed range of delay values over which a second-order delay plant can be stabilized by a PID, and more generally, a finite-dimensional LTI controller. This is done by deriving lower bounds on the delay margin achievable by

PD controllers, whereas this margin constitutes a lower bound on that achievable by PID controllers. Analysis shows that in most cases our results can be significantly less conservative than the lower bounds obtained elsewhere using more sophisticated, general LTI controllers.

5.4. Frequency-sweeping techniques in delayed dynamics analysis

Participants: Arben Cela [ESIEE Paris], Xu-Guang Li [Northeastern University, China], Xu Li [Northeastern University, China], Jiang-Chian Li [Northeastern University, China], Zhi-Zhong Mao [Shenyang University], Silviu Niculescu, Lu Zhang [Shenyang University].

The stability of linear systems with multiple (incommensurate) delays is investigated in [20], by extending a recently proposed frequency-sweeping approach. First, we consider the case where only one delay parameter is free while the others are fixed. The complete stability w.r.t. the free delay parameter can be systematically investigated by proving an appropriate invariance property. Next, we propose an iterative frequency-sweeping approach to study the stability under any given multiple delays. Moreover, we may effectively analyze the asymptotic behavior of the critical imaginary roots (if any) w.r.t. each delay parameter, which provides a possibility for stabilizing the system through adjusting the delay parameters. The approach is simple (graphical test) and can be applied systematically to the stability analysis of linear systems including multiple delays. A deeper discussion on its implementation is also proposed. Finally, various numerical examples complete the presentation.

In most of the numerical examples of time-delay systems proposed in the literature, the number of unstable characteristic roots remains positive before and after a multiple critical imaginary root (CIR) appears (as the delay, seen as a parameter, increases). This fact may lead to some misunderstandings: (i) A multiple CIR may at most affect the instability degree; (ii) It cannot cause any stability reversals (stability transitions from instability to stability). As far as we know, whether the appearance of a multiple CIR can induce stability is still unclear (in fact, when a CIR generates a stability reversal has not been specifically investigated). In [19], we provide a finer analysis of stability reversals and some new insights into the classification: the link between the multiplicity of a CIR and the asymptotic behavior with the stabilizing effect. Based on these results, we present an example illustrating that a multiple CIR's asymptotic behavior is able to cause a stability reversal. To the best of the authors' knowledge, such an example is a novelty in the literature on time-delay systems.

The work [30] focuses on the stability property of a class of distributed delay systems with constant coefficients. More precisely, we will discuss deeper the stability analysis with respect to the delay parameter. Our approach will allow to give new insights in solving the so-called complete stability problem. There are three technical issues need to be studied: First, the detection of the critical zero roots; second, the analysis of the asymptotic behavior of such critical zero roots; third, the asymptotic behavior analysis of the critical imaginary roots with respect to the infinitely many critical delays. We extended our recently-established frequency-sweeping approach, with which these technical issues can be effectively solved. More precisely, the main contributions of this paper are as follows: (i) Proposing a method for the detection of the critical zero roots. (ii) Proposing an approach for the asymptotic behavior analysis of such critical zero roots. (iii) The invariance property for the critical imaginary roots can be proved. Based on these results, a procedure was proposed, with which the complete stability analysis of such systems was accomplished systematically. Moreover, the procedure represents a unified approach: Most of the steps required by the complete stability problem may be fulfilled through observing the frequency-sweeping curves. Finally, some examples illustrate the effectiveness and advantages of the approach.

5.5. Weierstrass approach to asymptotic behavior of retarded differential equations

Participants: Jie Chen [City University of Hong Kong], Liliana Felix [University of San Luis Potosi], Alejandro Martinez-Gonzalez, Cesar F. Mendez-Barrios [University of San Luis Potosi], Silviu Niculescu.

The work [22] focuses on the analysis of the behavior of characteristic roots of time-delay systems, when the delay is subject to small parameter variations. The analysis is performed by means of the Weierstrass polynomial. More specifically, such a polynomial is employed to study the stability behavior of the characteristic roots with respect to small variations on the delay parameter. Analytic and splitting properties of the Puiseux series expansions of critical roots are characterized by allowing a full description of the cases that can be encountered. Several numerical examples encountered in the control literature are considered to illustrate the effectiveness of the proposed approach.

5.6. Some remarks on the Walton and Marshall method for neutral delay systems

Participants: Catherine Bonnet, Islam Boussaada, Le Ha Vy Nguyen.

The Walton and Marshall method allows to determine stability windows of delay systems of the retarded and neutral type. We noticed that some delay systems of the neutral type do not behave as claimed in [59] and analyzed carefully the position of the poles of such systems in the right half-plane. We have considered delay systems with characteristic equation being a quasi-polynomial with one delay and polynomials of degree one. It is shown that for a subclass of systems which have a chain of poles clustering the imaginary axis by the left, the procedure of Walton and Marshall fails: we prove the existence, for an infinitesimally small delay, of a positive real pole at infinity. This real pole is then proved to be the unique pole of the system in the closed right half-plane for all values of the delay. Some numerical examples illustrate the results [39]. We are currenlty extending those results to the general case of polynomials of degree $n > 1$.

5.7. $L2$ and BIBO Stability of systems with variable delays

Participants: Catherine Bonnet, Jonathan R. Partington [Univ of Leeds,U.K.].

There has been a growing interest in systems with time-varying delays in recent years and to the best of our knowledge, no study has been made of the input–output stability of a standard feedback scheme involving a system with time-varying delays. We have considered two versions of input–output stability: so-called Hinfinity and BIBO-stability. Retarded and neutral type systems with pointwise or distributed delays are studied here. We perform a stability analysis and discuss feedback stabilization. The results we derive yield very explicit estimates for stability margins [40].

5.8. Stability Analysis of Linear Partial Differential Equations

Participants: Giorgio Valmorbida, Aditya Gahlawat [University of Illinois at Urbana-Champaign].

We proposed a method to perform stability analysis of one-dimensional Partial Integro-Differential Equations. The relevance of the proposed results lies on the fact that we cast the Lyapunov inequalities as a differential inequality in two dimensions. The proposed structure for the inequalities is motivated by the same structure as the one used in the study of backstepping feedback laws, a successful strategy applied for several one-dimensional PDE systems. The advantage of the proposed Lyapunov analysis can be studied in a simpler manner as well as the fact that the backstepping law can be approximated by simpler laws and the stability can still be studied trhough the solution to the set of proposed inequalities.

We rely on Lyapunov analysis to establish the exponential stability of the systems. Then we present a test for the verification of the underlying Lyapunov inequalities, which relies on the existence of solutions of a system of coupled differential equations.

We illustrate the application of this method in several examples of PDEs defined by polynomial data, we formulate a numerical methodology in the form of a convex optimization problem which can be solved algorithmically. We show the effectiveness of the proposed numerical methodology using examples of different types of PDEs.

We are currently studying the extensions of coupled PDE-ODE systems.

5.9. Stability analysis of Piece-Wise Affine Systems

Participants: Giorgio Valmorbida, Leonardo Broering Groff, Joao Manoel Gomes Da Silva Jr [Universidade Federal do Rio Grande do Sul].

Piece-wise affine systems appear when linear dynamics are defined in different partitions of the state space. This type of system naturally appears whenever actuators have different stages or saturate or whenever non-linear control laws are obtained as the solution to a parameterised optimization problem as, for instance for systems with feedback laws based on the so-called explicit Model Predictive Control. Even though the dynamics is simple to describe, the stability analysis, performance assessment and robustness analysis are difficult to perform since, due to the often used explicit representation, the Lyapunov stability and dissipation tests are often described in terms of a number of inequalities that increase exponentially on the number of sets in the partition. Moreover regional stability and uncertainties corresponding to modification on the partition are difficult to study in this scenario.

To overcome these difficulties we have proposed an implicit representation for this class of systems in terms of ramp functions. The main advantage of such a representation lies on the fact that the ramp function can be exactly characterized in terms of linear inequalities and a quadratic equation, namely a linear complementarity condition. Thanks to the characterization of the ramp function and the implicit description of the PWA system the verification of Lyapunov inequalities related to piecewise quadratic functions can be cast as a pair of linear matrix inequalities.

The obtained formulation opens several possibilities to study the class of piecewise affine systems and their robustness properties. Indeed the fact that some partitions are uncertain is more easily coped with the proposed approach as they are described as parametric uncertainties of the implicit representation. Also, systems of larger dimension can be studied.

The stability analysis of the particular subclass of systems given by asymmetric saturation can also be performed with discontinuous Lyapunov functions for discrete-time systems.

5.10. Set Invariance for Descriptor Systems

Participants: Giorgio Valmorbida, Ye Wang [Universidade Federal do Rio Grande do Sul], Sorin Olaru [L2S, CentraleSupélec], Vicenç Puig [Universitat Politècnica de Catalunya], Gabriela Cembrano [Universitat Politècnica de Catalunya].

Some descriptor systems, characterized by a difference-algebraic relation, appear in economic systems. For such a class of systems we study robust invariant set characterizations of discrete-time descriptor systems and propose an active mode detection mechanism. The considered class of descriptor systems assumes regularity, stability and is affected by unknown-but-bounded disturbances.

As a first theoretical result, we establish a general framework for robust invariant sets for discrete-time descriptor systems in both causal and non-causal cases. Particular transformations are subsequently proposed for handling causal and non-causal descriptor systems and will be used to characterize the effects of disturbances. Based on these set-theoretic notions and a designed input signal for active set separations, we propose an active mode detection mechanism by exploiting the strong invariance properties.

5.11. New advances on backstepping

Participants: Frédéric Mazenc, Michael Malisoff [LSU], Laurent Burlion [Rutgers Univ.].

We worked on the problem of improving a major control design technique for nonlinear continuous-time systems called backstepping by using a fundamentally new approach which uses as key ingredient the introduction in the control of artificial delays or the use of dynamic extensions.

In the paper [24], we adopted a technique which is based on the introduction of pointwise delays (and not of distributed terms) to solve a challenging input-to-state stabilization problem for a chain of saturated integrators when the variables are not accurately measured. Let us observe that classical backstepping does not apply in the considered case.

In the paper [23], we constructed bounded globally asymptotically stabilizing output feedbacks for a family of nonlinear systems, using a dynamic extension and a converging-input-converging-state assumption. We provided sufficient conditions for this assumption to hold, in terms of Lyapunov functions. The novelty is that our construction provides formulas for the control bounds while allowing uncertainties that prevent the use of classical backstepping in cases where only part of the state variable is available. We illustrated our work with an engineering application: the single-link direct-drive manipulator.

5.12. Finite time observers

Participants: Frédéric Mazenc, Michael Malisoff [LSU], Saeed Ahmed [University of Kaiserslautern, Germany], Thach Dinh [CNAM], Tarek Raissi [CNAM].

Finite time observers are remarkably efficient when the value of the state is needed in a short time. By contrast with them, the solutions asymptotic observers may take a long time to be close to the solutions of the studied systems and may exhibit large transient errors. Motivated by this general fact, we produced three works.

In the work [12], we proposed finite time observers for time-varying nonlinear systems with delays in the outputs. When disturbances are present, we provide approximate values for the solutions which are expressed as upper bounds on the approximation errors after a suitable finite time. We illustrated our work via systems arising in the study of vibrating membranes, where time-varying coefficients can be used to represent intermittent measurements.

In paper [36], we use finite time reduced order continuous-discrete observers to solve an output feedback stabilization problem for a broad class of nonlinear systems whose output contains uncertainty. Unlike earlier works, our feedback control is discontinuous, but it does not contain any distributed terms, which is an advantage because the implementation of these terms may cause instability. We illustrated our main result by applying it to design a dynamic output feedback to solve a tracking problem for nonholonomic systems in chained form.

The paper [35] is devoted to the construction of finite time observers for discrete-time systems. We developped a new technique, which uses past values of the output. We considered the case where the systems are affected by additive disturbances and disturbances in the output. Exact estimation or approximate estimation have been achieved, depending on the absence or the presence of unknown but bounded uncertainties, respectively.

5.13. Observers with discrete measurements

Participants: Frédéric Mazenc, Michael Malisoff [LSU], Saeed Ahmed [University of Kaiserslautern, Germany].

We studied the important case where the measurements of a system are discrete because output of this type may preclude the use of observers designed under the assumption that the measurements are continuous.

In the paper [25], we have studied time-varying linear systems in the difficult case where the inputs and outputs have sampling and delays, and where the systems and outputs contain uncertainties. The observers we have proposed are of continuous-discrete type and have no distributed terms. We allowed the delays to be arbitrarily large and proved that the observer in combination with a linear control result in an input-to-state stability, under delays and sampling. We illustrated our work in two examples including a DC motor model.

In the work [38], we have revisited a well-known contribution of observer design for continuous-time systems with discrete measurements which relies on a dynamic extension. Using a stability analysis which relies on the recent technique called "trajectory based approach", we proved that, for systems with asynchronous sampling, the proposed dynamic observer is converging even when the size of some (sufficiently scarce) intervals between 2 measurements is larger than the upper bound ensuring convergence of the observer that is provided in the literature. A scarcety condition on these intervals is exhibited.

5.14. Observer for a rigid body

Participants: Frédéric Mazenc, Maruthi Akella [University of Texas, USA], Hongyang Duong [Harbin Institute of Technology], Qinglei Hu [Beihang University].

In the work [15], observers are proposed for models of rigid bodies in the case where only a part of the state variables can be measured. The design is based on the dual-quaternion description. To achieve tracking control objectives, the proposed observer are combined with an independently designed proportional-derivative-like feedback control law (using full-state feedback), and a special Lyapunov "strictification" process is employed to ensure a separation property between the observer and the controller. Almost global asymptotic stability of the closed-loop dynamics is guaranteed. We performed numerical simulations for a prototypical spacecraft pose tracking mission application to illustrate the effectiveness and robustness of the proposed method.

5.15. Systems with pointwise delays

Participants: Frédéric Mazenc, Michael Malisoff [LSU], Robledo Gonzalo [Univ. de Chile, Chile], Silviu Niculescu.

Frequently, the presence of delays is an obstacle to the stability analysis or the control of systems. Two of our works originate in the will to overcome this obstacle, in two distinct contexts.

The contribution [26] is devoted to the study of a model of a chain of two bio-reactors called 'chemostats'. One contains two microbial species in competition for a single limiting nutrient and receives an external input of the less advantaged competitor, which is cultivated in the other one. Pointwise delays are present. Under a condition on their size, we obtained sufficient conditions ensuring coexistence of all the species in competition. To prove the result, we adopted a Lyapunov based technique.

The contribution of [37] is twofold. In a first part, we exhibited a fundamental feature of the systems with a pointwise periodic time-varying delay: we have shown by a counterexample that the asymptotic stability of such a system cannot be deduced from the average value of the delay (even when the delay is 'rapidly' varying). In a second part, motivated by this counterexample, we proposed a new representation of systems with time-varying delays, which is helpful to carry out stability analyses and to develop a new state feedback stabilization method.

5.16. Control of an aircraft

Participants: Frédéric Mazenc, Michael Malisoff [LSU], Laurent Burlion [Rutgers Univ.].

The work [34] adapts some ideas we published in previous contributions to a control problem for a chain of saturating integrators for dynamics with outputs that occurs in the vision based landing of aircraft. The major obstacle that we overcome is due to the fact that only imprecise output measurements are available. The control laws we designed are bounded by an arbitrarily small constant. The closed-loop systems we obtained possess the robustness property called "local input-to-state stability".

5.17. Effect of environment on population dynamics

Participants: Islam Boussaada, Silviu Niculescu, Jun-Xiu Chen [Nothestern University], Sette Diop [L2S, CentraleSupélec], Xu-Guang Li [Nothestern University], Paul Raynaud de Fitte [LMRS, Université de Rouen Normandie], Safia Slimani [LMRS, Université de Rouen Normandie], Sami Tliba [L2S, CentraleSupélec].

Competition in population dynamics is often considered to be governed by predator-prey models. In particular, Lotka-Volterra models are intensively used in this context.

A modified version of a prey-predator system with Leslie-Gower and Holling type II functional responses incorporating a refuge for preys is studied in [27]. Such a refuge substantially complicates the dynamics of the system. We study the local and global dynamics and the existence of limit-cycles. We also investigate conditions for extinction or existence of a stationary distribution, in the case of a stochastic perturbation of the system.

Most of the reported Lotka-Volterra examples have at most one stability interval for the delay parameters. Furthermore, the existing methods fall short in treating more general case studies. Inspired by some recent results for analyzing the stability of time-delay systems, this paper focuses on a deeper characterization of the stability of Lotka-Volterra systems w.r.t. the delay parameters. In [58], we introduced the recently-proposed frequency-sweeping approach to study the complete stability problem for a broad class of linearized Lotka-Volterra systems. As a result, the whole stability delay-set is analytically determined. Moreover, as a significant byproduct of the proposed approach, some Lotka-Volterra examples are found to have multiple stability delay-intervals. In some situations, a longer maturation period of species is helpful for the stability of a population system.

In another context, the effect of environment on yeast population dynamics is studied in [28]. In presence of oxygen cells usually adopt efficient metabolism in order to maximize energy production yield in poor diet. If nutrient resource increases, a metabolic shift from efficient metabolism (respiration) to inefficient metabolism (fermentation) is reflecting a minimal cost principle of living systems to optimize fitness. This is known as the Crabtree/Warburg effect. A model that describes the population dynamics of cells and the input growth condition is established. Proof of principle has been constructed using a battery of growth experiments on Crabtree-positive yeasts–Saccharomyces under various conditions of glucose in aerobic and micro-aerobic conditions. General cell growth model estimating metabolic shift has been constructed based on an Auto-Regressive approach.

5.18. Distributed Algorithms for Microbiological systems

Participants: Lotfi Baour, Catherine Bonnet, Da-Jung Cho, Walid Djema [BIOCORE project team], Lucas Leclerc, Matthias Fuegger [MEXICO project team], Frédéric Mazenc, Tomas Nowak [LRI, Univ Paris-Sud], Cristina Stoica [CentraleSupélec].

This project led by M. Fuegger and T. Nowak considers microbiological distributed systems such as bacterial cultures. Our goal is to develop algorithms to be used in such low-powered biological computing systems. We have considered a distributed computation model where bacteria communicate via concentrations of small signaling molecules. Algorithms designed for such a model are of importance to provide services like clock synchronization. Basic ODE models for well-mixed solutions are readily available from systems biology. This year, we have paid attention to ODE and PDE models presented in [56], [55]. We have analyzed the influence of model parameters on the stability of the system. Moreover, looking at the involved basic biochemical reactions we have considered some changes in the modelling.

6. Partnerships and Cooperations

6.1. Regional Initiatives

Islam Boussaada is a deputy director of the IRS iCODE Institute, the institute for control and decision of the Idex Paris Saclay (http://icode-institute.fr).

- The project *Distributed Algorithms for Microbiological Systems* was funded by iCODE.
- The project *Symbolic/Numerical Methods and Implementations in Delayed-Control design* was funded by iCODE.
- The project *From modeling to control of microalgae growth in photo-bioreactor* was funded by iCODE.
- The project *Distributed Algorithms for Microbiological systems* was funded by iCODE.

6.2. National Initiatives

Islam Boussaada is a member of the administration council of the Association SAGIP (https://www.sagip.org), which structures and promotes the disciplines of automatic control and industrial engineering at the national level.

6.2.1. ANR

Giorgio Valmorbida is a member of the ANR HANDY - Hybrid And Networked Dynamical sYstems (http://projects.laas.fr/handy). Project Summary: Networked dynamical systems are ubiquitous in current and emerging technologies. From energy grids, fleets of connected autonomous vehicles to online social networks, the same scenario arises in each case: dynamical units interact locally to achieve a global behavior. When considering a networked system as a whole, very often continuous-time dynamics are affected by instantaneous changes, called jumps, leading to so-called hybrid dynamical systems. The jumps may originate from (i) the intrinsic dynamics of the nodes, like in multimedia delivery with fixed rate encoding, (ii) intrinsic limitations of the control actions, possibly constrained to a finite set of possible selections, like in power converters within energy grids, (iii) the creation/loss of links or the addition/removal of nodes like in renewable energy systems and social networks. Hybrid phenomena thus play an essential role in these control applications, and call upon the development of novel adapted tools for stability and performance analysis and control design. In this context, the aim of HANDY project is to provide methodological control-oriented tools for realistic networked models, which account for hybrid phenomena.

6.3. European Initiatives

6.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: **COST Action**

Project acronym: FRACTAL

Project title: Fractional-order systems; analysis, synthesis and their importance for future design

Duration: November 2016 - October 2020

Coordinator: Jaroslav Koton Czech Republic

Abstract: Fractional-order systems have lately been attracting significant attention and gaining more acceptance as generalization to classical integer-order systems. Mathematical basics of fractional-order calculus were laid nearly 300 years ago and since that it has gained deeply rooted mathematical concepts. Today, it is known that many real dynamic systems cannot be described by a system of simple differential equation or of integer-order system. In practice we can encounter such systems in electronics, signal processing, thermodynamics, biology, medicine, control theory, etc. The Action will favor scientific advancement in above mentioned areas by coordinating activities of academic research groups towards an efficient deployment of fractal theory to industry applications.

Program: **PHC AURORA**

Project acronym: -

Project title: Control and Observation of Nonlinear Systems

Duration: 01/2019-12/2019

Coordinator: Giorgio VALMORBIDA

Other partners: NTNU, Norvège

Abstract: Control theory and controller design for linear dynamical systems is well developed. The same cannot be said for nonlinear systems and searching for a general set of design tools applicable to any nonlinear system may be futile. Restricting the class of system dynamics with the aim of developing a more complete set of controller design tools for such a restricted model class therefore appears to be a reasonable approach. One such restricted class of system dynamics is the class of polynomial dynamical systems, for which stability analysis and controller design tools based on Convex Optimization has recently flourished, using so-called Sum of Squares (SOS) programming. Three topics were studied: - Time discretization techniques. SOS programming for discrete time system is less developed than for continuous time systems. This research task will then study discretization techniques leading to polynomial or rational models. In particular we will develop methods to compare the continuous time system and the discretized one by, for instance, comparing estimates of the region of attraction of stable equilibria. - Observer design. In many applications, not all states are measured, and therefore they have to be inferred using a state observer. Note that the so-called Certainty Equivalence Principle does not in general hold for nonlinear systems. This research task will therefore address observer design using SOS programming, and study the effects of interactions between controller design and observer design on the stability of the overall system. - Benchmark application. CentraleSupelec has a cart and pendulum experimental setup. The complexity of SOS-based controller design for this system is near the limit of what can be accommodated by current optimization packages and computational resources. This research task will test the limits of available numerical solution tools and provide a convincing demonstration of the capabilities of SOS-based controller design.

Program: **PHC BALATON**

Project acronym: SadHuB

Project title: Analysis of stabilizability of delayed dynamical system as function of the systems parameters and the time delays with applications to human balancing

Duration: 01/2018-12/2019

Coordinator: Islam Boussaada

Other partners: Budapest University of Technology and Economics, Hungary

Abstract: Motivated by a class of Time-delay systems occurring in modeling of many mechanical engineering applications, this project aims to associate researchers from control theory, applied mathematics and mechanical engineering to build together a general methodology for the analysis and control of mechanical/bio-mechanical structures. In particular, the human balance is often considered as a control system which operates in the presence of delays, primarily due to the time it takes to acquire the information needed for decision-making, to create control decisions, and to execute these decisions. A particular interest will be devoted to the delayed human balance, where a depthful study of the delay effect on the stability is expected.

6.4. International Initiatives

6.4.1. Inria International Partners

6.4.1.1. Informal International Partners

- Louisiana State University, Baton Rouge, USA
- Rutgers University, USA
- CINVESTAV, IPN, Mexico-City, Mexico
- Southern Illinois University, USA
- The University of Texas at Austin, Dept. of Aerospace Engineering & Engineering Mechanics, USA
- City University of Hong Kong, China
- Czech technical university in Prague, Czech Republic
- Budapest University of Technology and Economics, Hungary
- Katholieke Universiteit Leuven, Belgium
- Blikent University, Turkey
- Northeastern University, China
- Northeastern University, Boston, USA
- Universidad de Chile, Chile
- School of Mathematics, University of Leeds, U.K.
- UNICAMP, Brazil

- Kyoto University, Japan
- University Badji Mokhtar-Annaba, Algeria
- University Mouloud-Mammeri Tizi Ouzou, Algeria
- Universitat Politècnica de Catalunya, Spain
- University of Melbourne, Australia

6.4.2. Participation in Other International Programs

The team is member of the GDRI (International Research Group funded by CNRS) SpaDisco (following the GDRI Delsys) since 2017.

6.5. International Research Visitors

6.5.1. Visits of International Scientists

Jie Chen, CityU Hong Kong, 16-20 Dec 2019.

André Fioravanti, UNICAMP, Brazil, 1-7 Dec 2019.

Dan Ma, Northeastern University, 16-20 Dec 2019.

Hitay Özbay, Bilkent University, 4-8 Dec 2019.

Matheus Souza, UNICAMP, Brazil, 1-7 Dec 2019.

Joao Manoel Gomes da Silva Jr, UFRGS, Brazil, 15 Jul -15 Ago 2019.

Ross Drummond, University of Oxford, U.K., 1-7 Dec 2019.

Yutaka Yamamoto, Kyoto University, Japan, 30 oct - 8 Dec 2019.

6.5.1.1. Internships

Master internship: Lotfi Baour, Qualitative behaviour of two models of bacteria communication, Université de Cergy-Pontoise. Supervisors: Catherine Bonnet, Walid Djema, Matthias Fuegger and Thomas Nowak.

Master internship: Khaoula El Farhani, Modeling, estimation and control of microalgae growth for energy production and synthesis of molecules of high added values, CentraleSupelec, 05-09/2019. Supervisors: Sette Diop and Islam Boussaada.

Master internship: Jawher Kahouli, estimation and modelling of microalgae growth in photobioreactor, IPSA/Sup'Biotech,02-08/2019. Supervisors: Islam Boussaada, Ali El Ati and Jean-Yves Trosset.

Master internship: Robin Lacombe, qualification and start-up of Synoxis nano 2l photobioreactor, IPSA/Sup'Biotech,02-08/2019. Supervisors: Islam Boussaada, Ali El Ati and Jean-Yves Trosset.

Master internship: Lucas Leclerc, Modelling of bacteria communication through EDO/EDP, CentraleSupélec. Supervisor: Catherine Bonnet, Matthias Fuegger and Thomas Nowak.

Master internship: Javier Eduardo Pereyra Zamundio, New backstepping design using satificial delays for systems with pointwise delays, CINVESTAV, Instituto Politecnico Nacional. Supervisors: Sabine Mondié, Frédéric Mazenc.

6.5.2. Visits to International Teams

Islam Boussaada visited Budapest University of Technology and Economics during 1-7 Dec 2019.

Giorgio Valmorbida visited the University of Oxford 15-17 Jul 2019.

Giorgio Valmorbida visited the UFRGS, CEFET-Divinopolis, and the UNICAMP, Brazil 26 Jul - 6 Ago 2019.

7. Dissemination

7.1. Promoting Scientific Activities

7.1.1. Scientific Events: Organisation

7.1.1.1. General Chair, Scientific Chair

Islam Boussaada is co-leading the national research group *Tools for analysis and synthesis of infinite-dimensional systems* (GT OSYDI) of the CNRS/GDR MACS (https://gdr-macs.cnrs.fr/groupe-de-travail/osydi).

7.1.1.2. Member of the Organizing Committees

Catherine Bonnet was member of the organizing committee of SIAM CT19, Chendu China, July 2019.

7.1.2. Scientific Events: Selection

- Catherine Bonnet was Associate Editor for the 2019 American Control Conference, Philadelphia, USA and the 2020 American Control Conference, Denver, USA.
- Catherine Bonnet and Islam Boussaada were Associate Editor for the Joint IFAC Conference 7th Symposium on System Structure and Control (SSSC 2019) and 15th IFAC Workshop on Delay Systems, Sinaia, Roumania, Sept 2019.
- Frédéric Mazenc was Associate Editor for the conferences 2019 American Control Conference, Philadelphia, USA and for the European Control Conference 2020, St Petersburg.
- Giorgio Valmorbida was an Associate Editor for the 2019 American Control Conference, Philadelphia, USA and the 2020 American Control Conference, Denver, USA. He was also Associate editor for the 58th Conference in Decision and Control, Nice, France. Silviu Niculescu is the IEEE CSS (Control System Society) representative for the series of IEEE Conferences (nternational Conference on System Theory, Control and Computing,): http://icstcc2019.cs.upt.ro

7.1.2.1. Member of the Conference Program Committees

- Catherine Bonnet and Giorgio Valmorbida are members of the scientific committee of the GDRI (International Research Group funded by CNRS) SpaDisco since 2017.
- Giorgio Valmorbida is a member of the steering committee of the GDRI (International Research Group funded by CNRS) SpaDisco since 2017.
- Catherine Bonnet and Islam Boussaada were members of the International Program Committee of the Joint IFAC Conference 7th Symposium on System Structure and Control (SSSC 2019) and 15th IFAC Workshop on Delay Systems, Sinaia, Roumania, Sept 2019.

7.1.2.2. Reviewer

The team reviewed papers for several international conferences including IEEE Conference on Decision and Control, IEEE American Control Conference, European Control Conference ...

7.1.3. Journal

7.1.3.1. Member of the Editorial Boards

Frédéric Mazenc is member of the editorial boards of the following journals:

- Asian Journal of Control : Editor;
- IEEE Transactions on Automatic Control, Associate Editor;
- European Journal of Control, Associate Editor;
- Journal of Control and Decision, Associate Editor;
- IEEE Control Systems Letters, Associate Editor.

Silviu Niculescu is the Founding Editor and the Editor-in-Chief (since its creation in 2012) of the Springer (hard-cover) book series Advances in Delays and Dynamics ADD@S https://www.springer.com/series/11914.

He is Associate Editor at European Journal of Control (since 2011).

7.1.3.2. Reviewer - Reviewing Activities

The team reviewed papers for several journals including SIAM Journal on Control and Optimization, Automatica, IEEE Transactions on Automatic Control, Systems and Control Letters ...

7.1.4. Invited Talks

Giorgio Valmorbida was invited to give a talk in the workshop Scientific Computing Across Scales: Extreme Events and Criticality in Fluid Mechanics in April 15 - 18, 2019, at The Fields Institute 15-17, Toronto, CA.

Frédéric Mazenc was invited to give a talk in the CINVESTAV, IPN: Model Reduction and Predictor Control, in August 2, 2019, Mexico City, Mexico.

7.1.5. Leadership within the Scientific Community

Catherine Bonnet is a member of the IFAC Technical Committees on *Distributed Parameter Systems* and on *Biological and Medical Systems*. She is a member of the management committee of the COST Action FRACTAL (2016-2020).

Silviu Niculescu is the chair of the IFAC TC 2.2 "Linear Control Systems" since 2017 (including 300-350 researchers throught the world). The TC is coordinating 4 "Working Groups" (WG) including the WG on "Time-Delay Systems".

7.1.6. Scientific Expertise

Since September 2015, Catherine Bonnet is a member of the Evaluation Committee of Inria and since 2019 of the Bureau of the Evaluation Committee of Inria.

Since September 2016, Islam Boussaada is a member of the Scientific Council of IPSA (Engineering School in Aeronautic and Aerospace approved by CTI).

Since September 2018, Islam Boussaada is a member of the Development Council of Sup'Biotech (Engineering School in Biotechnologies approved by CTI).

Since 2014, Frédéric Mazenc is an expert for the FNRS (Belgium). His mission consists in evaluating research projects funded by this institution.

Since 2012, Frédéric Mazenc is a, expert for the ANVUR (National Agency for the Evaluation of Universities and Research Institutes, Italy). His mission consists in evaluating the contribution of Italian scientists.

Since 2011, Frédéric Mazenc is a, expert for the Romanian National Council for Development and Innovation (Romania). His mission consists in evaluating research projects funded by the this institution.

7.1.7. Research Administration

Catherine Bonnet is a member of the

- *Parity Committee* of Inria created sice its creation in 2015.
- Bureau du Comité des Projets du CRI Saclay-Ile-de-France since 2018.
- Coordination committee of the Mentoring Program of Inria Saclay-Île-de-France.
- PhD referent committee at L2S, CentraleSupelec.
- Administration council of the association *Femmes et Mathématiques*

Since October 2017, Frédéric Mazenc is Correspondant Inria Saclay A.M.I.E.S., http://www.agence-maths-entreprises.fr/. Since September 2019, he is member of "la Commission de Déloppement Technologique". He is member of the “Conseil du Laboratory of Signal and Systems".

7.2. Teaching - Supervision - Juries

7.2.1. Teaching

Licence: Islam Boussaada, *Control of bioprocesses*, 27h, 1st year, CentraleSupélec Université Paris-Saclay, France.

Licence: Silviu Niculescu, *Mathematics*, 15h, 1st year, ENSMP Paris, France.

Licence: Silviu Niculescu, *Introduction to optimization*, 30h, 1st year, ESIEE Paris, France.

Licence: Giorgio Valmorbida, *Signal Processing*, 1st year, 43h CentraleSupélec Université Paris-Saclay.

Master: Giorgio Valmorbida, *Tutorials - Modelling, Control, Optimization and Electrical Machines*, 1st and 2nd years, 55.5h, CentraleSupélec Université Paris-Saclay.

Master: Giorgio Valmorbida, *Projects and Internship supervision*, 2nd and 3rd years, 81h, CentraleSupélec Université Paris-Saclay.

Master: Giorgio Valmorbida, *Nonlinear Systems*, 3h, CentraleSupélec Executive Education, Université Paris-Saclay.

Master : Catherine Bonnet, *Stability properties and stabilization of interconnected dynamical systems involving delays*, 20h, IPSA, France.

Master : Silviu Niculescu, *Signals and Systems*, 12h, ESIEE Paris, France.

Master : Giorgio Valmorbida, *Control*, 40.5, Master MAE (M1), Université Paris-Saclay.

Master : Giorgio Valmorbida, *Stability of Dynamical Systems*, Master ATSI (M2), Université Paris-Saclay.

Doctorat : Silviu Niculescu, *Controlling Delayed Dynamics: Advances in Theory, Methods and Applications*, 7h, CISM Udine, Italy.

7.2.2. Supervision

PhD in progress: Souad Amrane, on real pole-placement for retarded functional differential equations, University Mouloud Mammeri. Since 09/2017. Supervisors: Fazia Bedouhene and Islam Boussaada.

PhD in progress: Amina Benarab, Characterization of the exponential decay of linear delay systems solutions, University Paris Saclay. Since 10/2019. Supervisors: Catherine Bonnet, Islam Boussaada and Karim Trabelsi.

PhD: Caetano Cardeliquio, Contributions to the Theory of Time-Delay Systems: Stability and Stabilisation, UNICAMP & Université Paris Saclay, 27 September 2019; Supervisors : Catherine Bonnet and André Fioravanti.

PhD in progress: Leonardo Broering Groff, Commande Périodique Déclenché par Événements, Université Paris-Saclay et UFRGS. Since 03/2017. Supervisors: Giorgio Valmorbida and Joao Manoel Gomes da Silva Jr.

PhD in progress: Jose Castillo, Design, Modeling and control of multi drones for aerial handling, University Paris Saclay, 10/2018, Islam Boussaada and Juan Escareno.

PhD in progress: Naouel Debiane, Bond Graph modeling for robust control and diagnosis of macatronic systems, University of Lille. Since 03/2017. Supervisors: Belkacem Ould-Bouamama and Islam Boussaada.

PhD in progress: Ali Diab, Commande par filtrage non linéaire des systèmes d'assistance direction, Université Paris-Saclay. Since 10/2019. Supervisors: Giorgio Valmorbida and William Pasillas-Lepine.

PhD in progress: Ricardo Falcon Prado, Active vibration control of flexible structures under input saturation through delay-based controllers and anti- windup compensators, University Paris Saclay. Since 10/2019. Supervisors: Islam Boussaada and Sami Tliba.

PhD in progress: Javier Eduardo Pereyra Zamundio, New backstepping design for systems with delay: finite time stabilization, robust stabilization, CINVESTAV, Instituto Politecnico Nacional. Since 10/2019. Supervisors: Sabine Mondié and Frédéric Mazenc.

PhD in progress : Amira Remadna, On pole-placement approach for retarded functional differential equations, University Badji Mokhtar-Annaba. Since 09/2019. Supervisors: Islam Boussaada and Azzedine Benchettah.

Postdoc: Da-Jung Cho, Modelling of bacteria communication, May-August 2019. Supervisors: Catherine Bonnet, Matthias Fuegger and Thomas Nowak.

7.2.3. Juries

- Catherine Bonnet was member of the Grenoble and Nancy Junior Researcher Inria recruiting committees.
- Catherine Bonnet was President of the PhD thesis of Yanqiao Wei *'Non-asymptotic and Robust fractional order differentiators using generalized modulating functions'*, 15 November 2019, INSA Val de Loire.

7.3. Popularization

7.3.1. Interventions

- The team welcomed Sophie Merheb, high-school student of Lycée privé Notre-Dame Les Oiseaux, Verneuil-sur-Seine, for the period 17-26 June 2019.

8. Bibliography

Major publications by the team in recent years

[1] C. BONNET, A. R. FIORAVANTI, J. R. PARTINGTON.*Stability of Neutral Systems with Commensurate Delays and Poles Asymptotic to the Imaginary Axis*, in "SIAM Journal on Control and Optimization", March 2011, vol. 49, n^{o} 2, p. 498-516, https://hal.inria.fr/hal-00782325

[2] C. BONNET, J. PARTINGTON.*Stabilization of some fractional delay systems of neutral type*, in "Automatica", 2007, vol. 43, p. 2047–2053

[3] I. BOUSSAADA, I.-C. MORARESCU, S.-I. NICULESCU.*Inverted pendulum stabilization: characterization of codimension-three triple zero bifurcation via multiple delayed proportional gains*, in "Systems Control Lett.", 2015, vol. 82, p. 1–9

[4] M. MALISOFF, F. MAZENC.*Constructions of Strict Lyapunov Functions*, Communications and Control Engineering Series, Springer-Verlag London Ltd., 2009

[5] F. MAZENC, M. MALISOFF, S.-I. NICULESCU.*Reduction Model Approach for Linear Time-Varying Systems with Delays*, in "IEEE Transactions on Automatic Control", 2014, vol. 59, n^{o} 8, p. 2068–2014

[6] W. MICHIELS, S.-I. NICULESCU. , S.-I. NICULESCU, W. MICHIELS (editors)*Stability, Control, and Computation for Time-Delay Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014

[7] S.-I. NICULESCU.*Delay Effects on Stability: a Robust Control Approach*, Lecture Notes in Control and Information Sciences, Springer, 2001, vol. 269

[8] M. B. SALDIVAR, I. BOUSSAADA, H. MOUNIER, S.-I. NICULESCU.*Analysis and Control of Oilwell Drilling Vibrations*, Springer International Publishing, May 2015, https://hal.archives-ouvertes.fr/hal-01123773

[9] G. VALMORBIDA, M. AHMADI, A. PAPACHRISTODOULOU.*Stability Analysis for a Class of Partial Differential Equations via Semidefinite Programming*, in "IEEE Transactions on Automatic Control", June 2016, vol. 61, n^{o} 6, p. 1649-1654

[10] G. VALMORBIDA, S. TARBOURIECH, G. GARCIA.*Design of Polynomial Control Laws for Polynomial Systems Subject to Actuator Saturation*, in "IEEE Transactions on Automatic Control", July 2013, vol. 58, n^{o} 7, p. 1758-1770

Publications of the year

Articles in International Peer-Reviewed Journal

[11] M. AHMADI, G. VALMÓRBIDA, D. GAYME, A. PAPACHRISTODOULOU.*A Framework for Input-Output Analysis of Wall-Bounded Shear Flows*, in "Journal of Fluid Mechanics", August 2019, vol. 873, p. 742-785 [*DOI :* 10.1017/JFM.2019.418], https://hal.archives-ouvertes.fr/hal-02430328

[12] S. AHMED, M. MALISOFF, F. MAZENC.*Finite Time Estimation for Time-Varying Systems with Delay in the Measurements* *, in "Systems and Control Letters", November 2019, https://hal.inria.fr/hal-02342854

[13] I. BOUSSAADA, S.-I. NICULESCU, A. EL ATI, R. PÉREZ-RAMOS, K. LIVIU TRABELSI.*Multiplicity-Induced-Dominancy in parametric second-order delay differential equations: Analysis and application in control design*, in "ESAIM: Control, Optimisation and Calculus of Variations", November 2019, forthcoming [*DOI :* 10.1051/COCV/2019073], https://hal.archives-ouvertes.fr/hal-02421253

[14] C. CARDELIQUIO, A. FIORAVANTI, C. BONNET, S.-I. NICULESCU.*Stability and Stabilisation Through Envelopes for Retarded and Neutral Time-Delay Systems*, in "IEEE Transactions on Automatic Control", 2019, 1, forthcoming [*DOI :* 10.1109/TAC.2019.2929092], https://hal-centralesupelec.archives-ouvertes.fr/hal-02350927

[15] H. DONG, Q. HU, M. R. AKELLA, F. MAZENC.*Partial Lyapunov Strictification: Dual Quaternion based Observer for 6-DOF Tracking Control*, in "IEEE Transactions on Control Systems Technology", November 2019 [*DOI :* 10.1109/TCST.2018.2864723], https://hal.inria.fr/hal-02342866

[16] A. GAHLAWAT, G. VALMÓRBIDA.*Stability Analysis of Linear Partial Differential Equations with Generalized Energy Functions*, in "IEEE Transactions on Automatic Control", July 2019 [*DOI :* 10.1109/TAC.2019.2927711], https://hal.archives-ouvertes.fr/hal-02430264

[17] D. IOAN, S. OLARU, I. PRODAN, F. STOICAN, S.-I. NICULESCU.*From Obstacle-Based Space Partitioning to Corridors and Path Planning. A Convex Lifting Approach*, in "IEEE Control Systems Letters", 2020, vol. 4, n^{o} 1, p. 79-84 [*DOI :* 10.1109/LCSYS.2019.2922414], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328848

[18] C. JIN, K. GU, I. BOUSSAADA, S.-I. NICULESCU.*Stability Analysis of a More General Class of Systems with Delay-Dependent Coefficients*, in "IEEE Transactions on Automatic Control", January 2019, vol. 14, 1 [*DOI :* 10.1109/TAC.2018.2832459], https://hal.archives-ouvertes.fr/hal-01957472

[19] X. Li, J.-C. Liu, X.-G. Li, S.-I. Niculescu, A. Cela.*Reversals in stability of linear time-delay systems A finer characterization*, in "Automatica", 2019, vol. 108, 108479 [*DOI :* 10.1016/j.automatica.2019.06.031], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328850

[20] X.-G. Li, S.-I. Niculescu, A. Cela.*An iterative frequency-sweeping approach for stability analysis of linear systems with multiple delays*, in "IMA Journal of Mathematical Control and Information", June 2019, vol. 36, n^{o} 02, p. 379-398 [*DOI :* 10.1093/imamci/dnx050], https://hal.archives-ouvertes.fr/hal-02281960

[21] D. Ma, J. Chen, A. Liu, J. Chen, S.-I. Niculescu.*Explicit bounds for guaranteed stabilization by PID control of second-order unstable delay systems*, in "Automatica", 2019, vol. 100, p. 407-411 [*DOI :* 10.1016/j.automatica.2018.11.053], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328846

[22] A. Martínez-González, C. F. Méndez-Barrios, S.-I. Niculescu, J. Chen, L. Félix.*Weierstrass approach to asymptotic behavior characterization of critical imaginary roots for retarded differential equations*, in "SIAM Journal on Control and Optimization", January 2019, vol. 57, n^{o} 1, p. 1-22 [*DOI :* 10.1137/17m1147135], https://hal.archives-ouvertes.fr/hal-01973973

[23] F. Mazenc, L. Burlion, M. Malisoff.*Backstepping Design for Output Feedback Stabilization for a Class of Uncertain Systems* *, in "Systems and Control Letters", January 2019 [*DOI :* 10.1016/j.sysconle.2018.11.007], https://hal.inria.fr/hal-02342856

[24] F. Mazenc, L. Burlion, M. Malisoff.*Stabilization and Robustness Analysis for a Chain of Saturating Integrators with Imprecise Measurements*, in "IEEE Control Systems Letters", April 2019, vol. 3, n^{o} 2, p. 428 - 433 [*DOI :* 10.1109/LCSYS.2019.2892931], https://hal.inria.fr/hal-02057620

[25] F. Mazenc, M. Malisoff.*Continuous-Discrete Sequential Observers for Time-Varying Systems under Sampling and Input Delays*, in "IEEE Transactions on Automatic Control", October 2019 [*DOI :* 10.1109/TAC.2019.2930676], https://hal.inria.fr/hal-02342849

[26] F. Mazenc, S.-I. Niculescu, G. Robledo.*Stability analysis of mathematical model of competition in a chain of chemostats in series with delay*, in "Applied Mathematical Modelling", 2019, vol. 76, p. 311-329 [*DOI :* 10.1016/j.apm.2019.06.006], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328851

[27] S. Slimani, P. Raynaud de Fitte, I. Boussaada.*Dynamics of a prey-predator system with modified Leslie-Gower and Holling type schemes incorporating a prey refuge*, in "Discrete & Continuous Dynamical Systems - B", 2019, vol. 24, n^{o} 9, p. 5003-5039 [*DOI :* 10.3934/dcdsb.2019042], https://hal.archives-ouvertes.fr/hal-02189210

[28] J.-Y. Trosset, S. Tliba, A. El Ati, H. Friha, E. Mogensen, J. Kahouli, R. Lacombe, O. Arkoun, S. Diop, R. Ledesma-Amaro, I. Boussaada.*Influence of micro-environment on yeast population dynamics*, in "Ecological Engineering and Environment Protection", 2019, vol. 2019, n^{o} 2, p. 5-13 [*DOI :* 10.32006/eeep.2019.2.0513], https://hal.archives-ouvertes.fr/hal-02413956

[29] Y. Wang, S. Olaru, G. Valmórbida, V. Puig, G. Cembrano.*Set-invariance characterizations of discrete-time descriptor systems with application to active mode detection*, in "Automatica", September 2019, vol. 107, p. 255-263 [*DOI :* 10.1016/j.automatica.2019.05.053], https://hal.archives-ouvertes.fr/hal-02430209

[30] L. ZHANG, Z.-Z. MAO, X.-G. LI, S.-I. NICULESCU, A. CELA.*Stability analysis for a class of distributed delay systems with constant coefficients by using a frequency-sweeping approach*, in "IET Control Theory and Applications", 2019, vol. 13, n^{o} 1, p. 87-95 [*DOI :* 10.1049/IET-CTA.2018.5520], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328849

Invited Conferences

[31] F. BEDOUHENE, I. BOUSSAADA, S.-I. NICULESCU.*On pole Placement and spectral abscissa characterization for time-delay systems*, in "7th Symposium on System Structure and Control (SSSC)", Sinaia, Romania, Islam Boussaada and Y Bouzidi and A Quadrat, September 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-02329054

[32] M. KUŘE, J. BUSEK, T. VYHLIDAL, S.-I. NICULESCU.*Damping oscillation of suspended payload by up and down motion of the pivot base time delay algorithms for UAV applications*, in "7th Symposium on System Structure and Control (SSSC) and 15th workshop on time-delay systems (TDS)", Sinaia, Romania, SSSC 2019 - Proceedings of joint IFAC 7th Symposium on System Structure and Control, Islam Boussaada and Y Bouzidi and A Quadrat, September 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-02329305

[33] S. TLIBA, I. BOUSSAADA, F. BEDOUHENE, S.-I. NICULESCU.*Active vibration control through quasi-polynomial based controller*, in "7th Symposium on System Structure and Control (SSSC) and 15th workshop on time-delay systems (TDS)", Sinaia, Romania, SSSC 2019 - Proceedings of joint IFAC 7th Symposium on System Structure and Control, Islam Boussaada and Y Bouzidi and A Quadrat, September 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-02329263

International Conferences with Proceedings

[34] L. BURLION, M. MALISOFF, F. MAZENC.*Stabilization and Robustness Analysis for a Chain of Saturating Integrators Arising in the Visual Landing of Aircraft*, in "CDC 2019 - 58th IEEE Conference on Decision and Control", Nice, France, December 2019, https://hal.inria.fr/hal-02342711

[35] T. N. DINH, F. MAZENC, T. RAÏSSI.*Finite-time guaranteed state estimation for discrete-time systems with disturbances*, in "SysTol 2019 - 4th International Conference on Control and Fault-Tolerant Systems", Casablanca, Morocco, IEEE, September 2019, p. 342-347 [*DOI :* 10.1109/SYSTOL.2019.8864742], https://hal.inria.fr/hal-02342708

[36] F. MAZENC, S. AHMED, M. MALISOFF.*Output Feedback Stabilization by Reduced Order Finite Time Observers using a Trajectory Based Approach*, in "ACC 2019 - American Control Conference", Philadelphia, United States, July 2019, https://hal.inria.fr/hal-02268093

[37] F. MAZENC, M. MALISOFF.*On Average Values of Time-Varying Delays and a New Representation of Systems with Time-Varying Delays*, in "ACC 2019 American Control Conference", Philadelphia, United States, July 2019, https://hal.inria.fr/hal-02268094

[38] F. MAZENC.*Sampled-Data Observers: Scarce Arbitrarly Large Sampling Intervals*, in "ICTSTCC 2019 - 23rd International Conference on System Theory, Control and Computing", Sinaia, Romania, October 2019, https://hal.inria.fr/hal-02342668

[39] V. NGUYEN, C. BONNET, I. BOUSSAADA, M. SOUAIBY.*A problematic issue in the Walton-Marshall method for some neutral delay systems*, in "58th Conference on Decision and Control - Nice, France - December 11th-13th 2019", Nice, France, December 2019, https://hal.inria.fr/hal-02431270

Conferences without Proceedings

[40] C. Bonnet, J. Partington. *L2 and BIBO-stability of systems with variable delays*, in "15th IFAC Workshop on Time Delay Systems", Sinaia, Romania, September 2019, https://hal.inria.fr/hal-02421180

[41] I. Boussaada, H. Mounier, S.-I. Niculescu. *Low complexity controllers for vibrations damping in drilling systems*, in "ACC 2019 - American Control Conference", Philadelphia, United States, Institute of Electrical and Electronics Engineers Inc., 2019, vol. 2019-July, p. 5228-5233, https://hal-centralesupelec.archives-ouvertes.fr/hal-02328847

[42] J. d. J. Castillo-Zamora, J. Escareño, J. Alvarez, J. Stephant, I. Boussaada. *Disturbances and Coupling Compensation for Trajectory Tracking of a Multi-link Aerial Robot*, in "CoDIT 2019 - 6th International Conference on Control, Decision and Information Technologies", Paris, France, IEEE, April 2019, p. 738-743 [*DOI :* 10.1109/CoDIT.2019.8820513], https://hal.archives-ouvertes.fr/hal-02417127

[43] L. B. Groff, J. M. Gomes Da Silva, G. Valmórbida. *Regional Stability of Discrete-Time Linear Systems Subject to Asymmetric Input Saturation*, in "58th IEEE Conference on Decision and Control", Nice, France, December 2019, https://hal.archives-ouvertes.fr/hal-02430391

[44] L. B. Groff, G. Valmórbida, J. M. Gomes Da Silva. *Stability analysis of piecewise affine discrete-time systems**, in "58th IEEE Conference on Decision and Control", Nice, France, December 2019, https://hal.archives-ouvertes.fr/hal-02430377

[45] J.-E. Hernández-Díez, C. F. Méndez-Barrios, S.-I. Niculescu, V. M. Ramirez Rivera. *Closed-loop stability analysis of voltage mode buck using a proportional-delayed-integral controller*, in "ICIEA 2019 - 14th IEEE Conference on Industrial Electronics and Applications", Xian, China, Institute of Electrical and Electronics Engineers Inc., 2019, p. 233-238 [*DOI :* 10.1109/ICIEA.2019.8834343], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328845

[46] D. Ioan, S. Olaru, S.-I. Niculescu, I. Prodan, F. Stoican. *Navigation in a multi-obstacle environment. from partition of the space to a zonotopic-based MPC*, in "ECC 2019 - 18th European Control Conference", Naples, Italy, Institute of Electrical and Electronics Engineers Inc., 2019, p. 1772-1777 [*DOI :* 10.23919/ECC.2019.8796080], https://hal-centralesupelec.archives-ouvertes.fr/hal-02328852

Scientific Books (or Scientific Book chapters)

[47] X.-G. Li, S.-I. Niculescu, A. Cela, L. Zhang. *Stability Analysis of Uniformly Distributed Delay Systems: A Frequency-Sweeping Approach*, in "Delays and Interconnections: Methodology, Algorithms and Applications", G. Valmorbida, I. Boussaada, A. Seuret, R. Sipahi (editors), Part of the Advances in Delays and Dynamics book series (ADVSDD), Springer, October 2019, vol. 10, p. 117-130 [*DOI :* 10.1007/978-3-030-11554-8_8], https://hal.archives-ouvertes.fr/hal-02350765

[48] A. Martínez-González, S.-I. Niculescu, J. Chen, C. F. Méndez-Barrios, J. G. Romero Velázquez, G. Mejia-Rodriguez. *Asymptotic Analysis of Multiple Characteristics Roots for Quasi-polynomials of Retarded-Type*, in "Delays and Interconnections: Methodology, Algorithms and Applications", G. Valmorbida, I. Boussaada, A. Seuret, R. Sipahi (editors), Part of the Advances in Delays and Dynamics book series (ADVSDD), Springer, October 2019, vol. 10, p. 131-151 [*DOI :* 10.1007/978-3-030-11554-8_9], https://hal.archives-ouvertes.fr/hal-02350881

[49] G. VALMORBIDA, A. SEURET, I. BOUSSAADA, R. SIPAHI.*Delays and Interconnections: Methodology, Algorithms and Applications*, Springer, September 2019 [*DOI :* 10.1007/978-3-030-11554-8], https://hal.laas.fr/hal-02346507

Research Reports

[50] D. IOAN, I. PRODAN, F. STOICAN, S. OLARU, S.-I. NICULESCU.*Complexity bounds for obstacle avoidance within a zonotopic framework*, L2S, CNRS, Supelec, Université Paris Sud, January 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-01986337

Other Publications

[51] Y. CHITOUR, G. MAZANTI, M. SIGALOTTI.*On the gap between deterministic and probabilistic joint spectral radii for discrete-time linear systems*, November 2019, https://arxiv.org/abs/1812.08399 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01961003

[52] G. MAZANTI, I. BOUSSAADA, S.-I. NICULESCU.*On qualitative properties of single-delay linear retarded differential equations: Characteristic roots of maximal multiplicity are necessarily dominant*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02422706

[53] G. MAZANTI, I. BOUSSAADA, S.-I. NICULESCU, T. VYHLÍDAL.*Spectral dominance of complex roots for single-delay linear equations*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02422707

[54] S. SLIMANI, P. RAYNAUD DE FITTE, I. BOUSSAADA.*Dynamics of a prey-predator system with modified Leslie-Gower and Holling type II schemes incorporating a prey refuge*, January 2019, https://arxiv.org/abs/1802.05967 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01707650

References in notes

[55] T. DANINO, O. MONDRAGÓN-PALOMINO, L. TSIMRING, J. HASTY.*A synchronized quorum of genetic clocks*, in "Nature", 2010, vol. 463, n° 7279, 326

[56] M. O. DIN, T. DANINO, A. PRINDLE, M. SKALAK, J. SELIMKHANOV, K. ALLEN, E. JULIO, E. ATOLIA, L. S. TSIMRING, S. N. BHATIA, ET AL. .*Synchronized cycles of bacterial lysis for in vivo delivery*, in "Nature", 2016, vol. 536, n° 7614, 81

[57] H. KITANO.*Biological robustness*, in "Nature", 2004, vol. 5, p. 826–837

[58] X.-G. LI, J.-X. CHEN, S.-I. NICULESCU, A. CELA.*New insights in stability analysis of delayed Lotka–Volterra systems*, in "Journal of The Franklin Institute", November 2018, vol. 355, n° 17, p. 8683-8697 [*DOI :* 10.1016/J.JFRANKLIN.2018.09.011], https://hal-centralesupelec.archives-ouvertes.fr/hal-02281715

[59] J. E. MARSHALL, H. GÓRECKI, A. KORYTOWSKI, K. WALTON.*Time-Delay Systems: stability and performance criteria with applications*, Ellis Horwood, 1992

[60] H. OZBAY, C. BONNET, A. FIORAVANTI.*PID Controller Design for Fractional-Order Systems with Time Delays*, in "Systems and Control Letters", January 2012, vol. 61, n° 1, p. 18-23, http://hal.inria.fr/hal-00766062

Project-Team EX-SITU

Extreme Interaction

IN COLLABORATION WITH: Laboratoire de recherche en informatique (LRI)

IN PARTNERSHIP WITH:
CNRS

Université Paris-Sud (Paris 11)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Interaction and visualization

Table of contents

Project-Team EX-SITU

Creation of the Team: 2015 January 01, updated into Project-Team: 2017 July 01

Keywords:

Computer Science and Digital Science:

A5.1. - Human-Computer Interaction
A5.1.1. - Engineering of interactive systems
A5.1.2. - Evaluation of interactive systems
A5.1.5. - Body-based interfaces
A5.1.6. - Tangible interfaces
A5.1.7. - Multimodal interfaces
A5.2. - Data visualization

Other Research Topics and Application Domains:

B2.8. - Sports, performance, motor skills
B5.7. - 3D printing
B6.3.1. - Web
B6.3.4. - Social Networks
B9.2. - Art
B9.2.1. - Music, sound
B9.2.4. - Theater
B9.5. - Sciences

1. Team, Visitors, External Collaborators

Research Scientists

Wendy Mackay [Team leader, Inria, Senior Researcher, HDR]
Theophanis Tsandilas [Inria, Researcher]
Baptiste Caramiaux [CNRS, Researcher]

Faculty Members

Joanna Mcgrenere [University of British Columbia and Inria Chair, Jun 2019]
Michel Beaudouin-Lafon [Univ Paris-Sud, Professor, HDR]
Sarah Fdili Alaoui [Univ Paris-Sud, Assistant Professor]
Cédric Fleury [Univ Paris-Sud, Assistant Professor]

External Collaborator

Midas Nouwens [Aarhus University]

Technical Staff

Lawrence Fyfe [Inria, Engineer, until Jan 2019]
Xi Hu [Univ Paris-Sud, Engineer, from Oct 2019 until Nov 2019]
Martin Tricaud [Univ Paris-Sud, Engineer, from Apr 2019 until Sep 2019]
Alexandre Battut [Univ Paris-Sud, Engineer, from Oct 2019]
Olivier Gladin [Inria, Engineer]
Nicolas Taffin [Inria, Engineer]

PhD Students

Dimitrios Christaras Papageorgiou [Univ Paris-Sud, PhD Student, until Nov 2019]

Per Carl Viktor Gustafsson [Univ Paris-Sud, PhD Student]
Han Han [Univ Paris-Sud, PhD Student]
Shu-Yuan Hsueh [Univ Paris-Sud, PhD Student]
Wanyu Liu [Institut Telecom, PhD Student, until Jan 2019]
Yujiro Okuya [Univ Paris-Sud, PhD Student, until Nov 2019]
Miguel Renom [Univ Paris-Sud, PhD Student]
Jean-Philippe Rivière [Univ Paris-Sud, PhD Student]
Téo Sanchez [Inria, PhD Student]
Philip Tchernavskij [Univ Paris-Saclay, PhD Student]
Martin Tricaud [CNRS, PhD Student, from Oct 2019]
Elizabeth Walton [Univ Paris-Sud, PhD Student]
Yi Zhang [Inria, PhD Student]
Yiran Zhang [Univ Paris-Sud, PhD Student]
Arthur Fages [Univ Paris-Sud, PhD Student, from Dec 2019]

Post-Doctoral Fellows

Benjamin Bressolette [Univ Paris-Sud, Post-Doctoral Fellow]
Julien Gori [Univ Paris-Sud, Post-Doctoral Fellow, from Feb 2019]
Carla Griggio [Inria, Post-Doctoral Fellow, until May 2019]

Visiting Scientists

Janin Koch [Aalto University, until Jan 2019]
Injung Lee [KAIST, from Oct 2019]

Administrative Assistants

Irina Lahaye [Inria, Administrative Assistant, from Nov 2019]
Alexandra Merlin [Inria, Administrative Assistant, until Sep 2019]

2. Overall Objectives

2.1. Overall Objectives

Interactive devices are everywhere: we wear them on our wrists and belts; we consult them from purses and pockets; we read them on the sofa and on the metro; we rely on them to control cars and appliances; and soon we will interact with them on living room walls and billboards in the city. Over the past 30 years, we have witnessed tremendous advances in both hardware and networking technology, which have revolutionized all aspects of our lives, not only business and industry, but also health, education and entertainment. Yet the ways in which we interact with these technologies remains mired in the 1980s. The graphical user interface (GUI), revolutionary at the time, has been pushed far past its limits. Originally designed to help secretaries perform administrative tasks in a work setting, the GUI is now applied to every kind of device, for every kind of setting. While this may make sense for novice users, it forces expert users to use frustratingly inefficient and idiosyncratic tools that are neither powerful nor incrementally learnable.

ExSitu explores the limits of interaction — how extreme users interact with technology in extreme situations. Rather than beginning with novice users and adding complexity, we begin with expert users who already face extreme interaction requirements. We are particularly interested in creative professionals, artists and designers who rewrite the rules as they create new works, and scientists who seek to understand complex phenomena through creative exploration of large quantities of data. Studying these advanced users today will not only help us to anticipate the routine tasks of tomorrow, but to advance our understanding of interaction itself. We seek to create effective human-computer partnerships, in which expert users control their interaction with technology. Our goal is to advance our understanding of interaction as a phenomenon, with a corresponding paradigm shift in how we design, implement and use interactive systems. We have already made significant progress through our work on instrumental interaction and co-adaptive systems, and we hope to extend these into a foundation for the design of all interactive technology.

3. Research Program

3.1. Research Program

We characterize Extreme Situated Interaction as follows:

Extreme users. We study extreme users who make extreme demands on current technology. We know that human beings take advantage of the laws of physics to find creative new uses for physical objects. However, this level of adaptability is severely limited when manipulating digital objects. Even so, we find that creative professionals—artists, designers and scientists—often adapt interactive technology in novel and unexpected ways and find creative solutions. By studying these users, we hope to not only address the specific problems they face, but also to identify the underlying principles that will help us to reinvent virtual tools. We seek to shift the paradigm of interactive software, to establish the laws of interaction that significantly empower users and allow them to control their digital environment.

Extreme situations. We develop extreme environments that push the limits of today's technology. We take as given that future developments will solve “practical" problems such as cost, reliability and performance and concentrate our efforts on interaction in and with such environments. This has been a successful strategy in the past: Personal computers only became prevalent after the invention of the desktop graphical user interface. Smartphones and tablets only became commercially successful after Apple cracked the problem of a usable touch-based interface for the iPhone and the iPad. Although wearable technologies, such as watches and glasses, are finally beginning to take off, we do not believe that they will create the major disruptions already caused by personal computers, smartphones and tablets. Instead, we believe that future disruptive technologies will include fully interactive paper and large interactive displays.

Our extensive experience with the Digiscope WILD and WILDER platforms places us in a unique position to understand the principles of distributed interaction that extreme environments call for. We expect to integrate, at a fundamental level, the collaborative capabilities that such environments afford. Indeed almost all of our activities in both the digital and the physical world take place within a complex web of human relationships. Current systems only support, at best, passive sharing of information, e.g., through the distribution of independent copies. Our goal is to support active collaboration, in which multiple users are actively engaged in the lifecycle of digital artifacts.

Extreme design. We explore novel approaches to the design of interactive systems, with particular emphasis on extreme users in extreme environments. Our goal is to empower creative professionals, allowing them to act as both designers and developers throughout the design process. Extreme design affects every stage, from requirements definition, to early prototyping and design exploration, to implementation, to adaptation and appropriation by end users. We hope to push the limits of participatory design to actively support creativity at all stages of the design lifecycle. Extreme design does not stop with purely digital artifacts. The advent of digital fabrication tools and FabLabs has significantly lowered the cost of making physical objects interactive. Creative professionals now create hybrid interactive objects that can be tuned to the user's needs. Integrating the design of physical objects into the software design process raises new challenges, with new methods and skills to support this form of extreme prototyping.

Our overall approach is to identify a small number of specific projects, organized around four themes: *Creativity, Augmentation, Collaboration* and *Infrastructure*. Specific projects may address multiple themes, and different members of the group work together to advance these different topics.

4. Application Domains

4.1. Creative industries

We work closely with creative professionals in the arts and in design, including music composers, musicians, and sound engineers; painters and illustrators; dancers and choreographers; theater groups; game designers; graphic and industrial designers; and architects.

4.2. Scientific research

We work with creative professionals in the sciences and engineering, including neuroscientists and doctors; programmers and statisticians; chemists and astrophysicists; and researchers in fluid mechanics.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Wanyu Liu: First prize, Télécom ParisTech thesis award, "Information theory as a unified tool for understanding and designing human-computer interaction"
- Stacy Hsueh, Sarah Fdili Alaoui, and Wendy Mackay: Honorable Mention Award at ACM CSCW 2019 for "Deconstructing Creativity: Non-Linear Processes and Fluid Roles in Contemporary Music and Dance." [21]
- Alexander Eiselmayer, Chat Wacharamanotham, Michel Beaudouin-Lafon, and Wendy Mackay: Best Paper award at ACM CHI 2019 for "Touchstone2: An Interactive Environment for Exploring Trade-offs in HCI Experiment Design" [19]

6. New Software and Platforms

6.1. Digiscape

KEYWORDS: 2D - 3D - Node.js - Unity 3D - Video stream

FUNCTIONAL DESCRIPTION: Through the Digiscape application, the users can connect to a remote workspace and share files, video and audio streams with other users. Application running on complex visualization platforms can be easily launched and synchronized.

- Partners: Maison de la simulation - UVSQ - CEA - ENS Cachan - LIMSI - LRI - Laboratoire de Recherche en Informatique - CentraleSupélec - Telecom Paris
- Contact: Olivier Gladin
- URL: http://www.digiscope.fr

6.2. Touchstone2

KEYWORD: Experimental design

FUNCTIONAL DESCRIPTION: Touchstone2 is a graphical user interface to create and compare experimental designs. It is based on a visual language: Each experiment consists of nested bricks that represent the overall design, blocking levels, independent variables, and their levels. Parameters such as variable names, counterbalancing strategy and trial duration are specified in the bricks and used to compute the minimum number of participants for a balanced design, account for learning effects, and estimate session length. An experiment summary appears below each brick assembly, documenting the design. Manipulating bricks immediately generates a corresponding trial table that shows the distribution of experiment conditions across participants. Trial tables are faceted by participant. Using brushing and fish-eye views, users can easily compare among participants and among designs on one screen, and examine their trade-offs.

Touchstone2 plots a power chart for each experiment in the workspace. Each power curve is a function of the number of participants, and thus increases monotonically. Dots on the curves denote numbers of participants for a balanced design. The pink area corresponds to a power less than the 0.8 criterion: the first dot above it indicates the minimum number of participants. To refine this estimate, users can choose among Cohen's three conventional effect sizes, directly enter a numerical effect size, or use a calculator to enter mean values for each treatment of the dependent variable (often from a pilot study).

Touchstone2 can export a design in a variety of formats, including JSON and XML for the trial table, and TSL, a language we have created to describe experimental designs. A command-line tool is provided to generate a trial table from a TSL description.

Touchstone2 runs in any modern Web browser and is also available as a standalone tool. It is used at ExSitu for the design of our experiments, and by other Universities and research centers worldwide. It is available under an Open Source licence at https://touchstone2.org.

- Partner: University of Zurich
- Contact: Wendy Mackay
- URL: https://touchstone2.org

6.3. UnityCluster

KEYWORDS: 3D - Virtual reality - 3D interaction

FUNCTIONAL DESCRIPTION: UnityCluster is middleware to distribute any Unity 3D (https://unity3d.com/) application on a cluster of computers that run in interactive rooms, such as our WILD and WILDER rooms, or immersive CAVES (Computer-Augmented Virtual Environments). Users can interact the the application with various interaction resources.

UnityCluster provides an easy solution for running existing Unity 3D applications on any display that requires a rendering cluster with several computers. UnityCluster is based on a master-slave architecture: The master computer runs the main application and the physical simulation as well as manages the input, the slave computers receive updates from the master and render small parts of the 3D scene. UnityCluster manages data distribution and synchronization among the computers to obtain a consistent image on the entire wall-sized display surface.

UnityCluster can also deform the displayed images according to the user's position in order to match the viewing frustum defined by the user's head and the four corners of the screens. This respects the motion parallax of the 3D scene, giving users a better sense of depth.

UnityCluster is composed of a set of C Sharp scripts that manage the network connection, data distribution, and the deformation of the viewing frustum. In order to distribute an existing application on the rendering cluster, all scripts must be embedded into a Unity package that is included in an existing Unity project.

- Partner: Inria
- Contact: Cédric Fleury

6.4. VideoClipper

KEYWORD: Video recording

FUNCTIONAL DESCRIPTION: VideoClipper is an IOS app for Apple Ipad, designed to guide the capture of video during a variety of prototyping activities, including video brainstorming, interviews, video prototyping and participatory design workshops. It relies heavily on Apple's AVFoundation, a framework that provides essential services for working with time-based audiovisual media on iOS (https://developer.apple.com/av-foundation/). Key uses include: transforming still images (title cards) into video tracks, composing video and audio tracks in memory to create a preview of the resulting video project and saving video files into the default Photo Album outside the application.

VideoClipper consists of four main screens: project list, project, capture and import. The project list screen shows a list with the most recent projects at the top and allows the user to quickly add, remove or clone (copy and paste) projects. The project screen includes a storyboard composed of storylines that can be added, cloned or deleted. Each storyline is composed of a single title card, followed by one or more video clips. Users can reorder storylines within the storyboard, and the elements within each storyline through direct manipulation. Users can preview the complete storyboard, including all titlecards and videos, by pressing the play button, or export it to the Ipad's Photo Album by pressing the action button.

VideoClipper offers multiple tools for editing titlecards and storylines. Tapping on the title card lets the user edit the foreground text, including font, size and color, change background color, add or edit text labels, including size, position, color, and add or edit images, both new pictures and existing ones. Users can also delete text labels and images with the trash button. Video clips are presented via a standard video player, with standard interaction. Users can tap on any clip in a storyline to: trim the clip with a non-destructive trimming tool, delete it with a trash button, open a capture screen by clicking on the camera icon, label the clip by clicking a colored label button, and display or hide the selected clip by toggling the eye icon.

VideoClipper is currently in beta test, and is used by students in two HCI classes at the Université Paris-Saclay, researchers in ExSitu as well as external researchers who use it for both teaching and research work. A beta test version is available on demand under the Apple testflight online service.

- Contact: Wendy Mackay

6.5. WildOS

KEYWORDS: Human Computer Interaction - Wall displays

FUNCTIONAL DESCRIPTION: WildOS is middleware to support applications running in an interactive room featuring various interaction resources, such as our WILD and WILDER rooms: a tiled wall display, a motion tracking system, tablets and smartphones, etc. The conceptual model of WildOS is a platform, such as the WILD or WILDER room, described as a set of devices and on which one or more applications can be run.

WildOS consists of a server running on a machine that has network access to all the machines involved in the platform, and a set of clients running on the various interaction resources, such as a display cluster or a tablet. Once WildOS is running, applications can be started and stopped and devices can be added to or removed from the platform.

WildOS relies on Web technologies, most notably Javascript and node.js, as well as node-webkit and HTML5. This makes it inherently portable (it is currently tested on Mac OS X and Linux). While applications can be developed only with these Web technologies, it is also possible to bridge to existing applications developed in other environments if they provide sufficient access for remote control. Sample applications include a web browser, an image viewer, a window manager, and the BrainTwister application developed in collaboration with neuroanatomists at NeuroSpin.

WildOS is used for several research projects at ExSitu and by other partners of the Digiscope project. It was also deployed on several of Google's interactive rooms in Mountain View, Dublin and Paris. It is available under an Open Source licence at https://bitbucket.org/mblinsitu/wildos.

- Contact: Michel Beaudouin-Lafon
- URL: https://bitbucket.org/mblinsitu/wildos

6.6. Platforms

6.6.1. WILDER

Participants: Michel Beaudouin-Lafon [correspondant], Cédric Fleury, Olivier Gladin.

WILDER (Figure 1) is our second experimental ultra-high-resolution interactive environment, which follows the WILD platform developed in 2009. It features a wall-sized display with seventy-five 20" LCD screens, i.e. a 5m50 x 1m80 (18' x 6') wall displaying 14 400 x 4 800 = 69 million pixels, powered by a 10-computer cluster and two front-end computers. The platform also features a camera-based motion tracking system that lets users interact with the wall, as well as the surrounding space, with various mobile devices. The display uses a multitouch frame (the largest of its kind in the world) to make the entire wall touch sensitive.

WILDER was inaugurated in June, 2015. It is one of the ten platforms of the Digiscope Equipment of Excellence and, in combination with WILD and the other Digiscope rooms, provides a unique experimental environment for collaborative interaction.

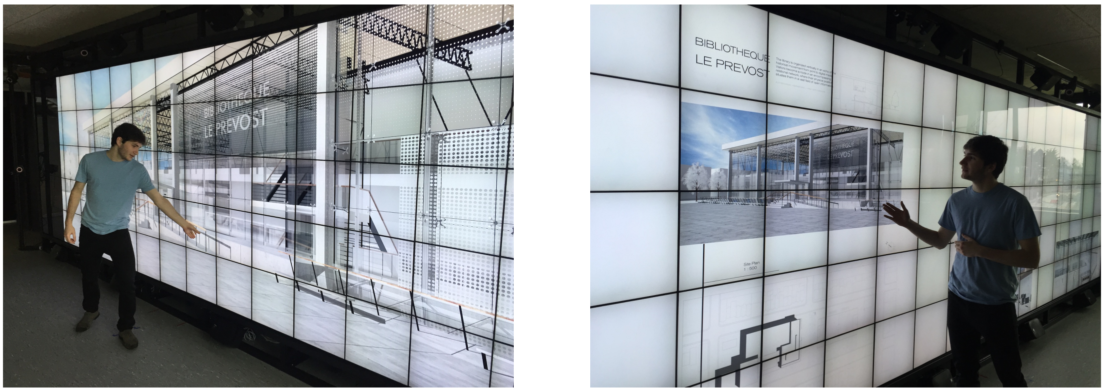


Figure 1. The WILDER platform.

In addition to using WILD and WILDER for our research, we have also developed software architectures and toolkits, such as WildOS and Unity Cluster, that enable developers to run applications on these multi-device, cluster-based systems.

7. New Results

7.1. Fundamentals of Interaction

Participants: Michel Beaudouin-Lafon [correspondant], Wendy Mackay, Cédric Fleury, Theophanis Tsandilas, Benjamin Bressolette, Julien Gori, Han Han, Yiran Zhang, Miguel Renom, Philip Tchernavskij, Martin Tricaud.

In order to better understand fundamental aspects of interaction, ExSitu conducts in-depth observational studies and controlled experiments which contribute to theories and frameworks that unify our findings and help us generate new, advanced interaction techniques. Our theoretical work also leads us to deepen or re-analyze existing theories and methodologies in order to gain new insights.

At the methodological level and in collaboration with University of Zurich (Switzerland), we have developed *Touchstone2* [19] (Best Paper award), a direct-manipulation interface for generating and examining trade-offs in experiment designs (Fig. 2). Based on interviews with experienced researchers, we developed an interactive environment for manipulating experiment design parameters, revealing patterns in trial tables, and estimating and comparing statistical power. We also developed TSL, a declarative language that precisely represents experiment designs. In two studies, experienced HCI researchers successfully used *Touchstone2* to evaluate design trade-offs and calculate how many participants are required for particular effect sizes. *Touchstone2* is freely available at https://touchstone2.org and we encourage the community to use it to improve the accountability and reproducibility of research by sharing TSL descriptions of their experimental designs.

The book "Sticky Creativity: Post-It Note Cognition, Interaction and Digitalization" [32], Academic Press, explores how the Post-It note has "become the most commonly used design material in creative design activities", with research and use cases to illustrate its role creative activities. Wendy Mackay converted her one-day Master Class on participatory design methods into a book chapter, shifting the designer's focus from static wireframes to prototyping how users will interact with a proposed new technology. The course takes the reader through a full interaction design cycle, with nine illustrated participatory design methods. It begins with a design brief: create an augmented sticky note inspired by observations of how people actually use

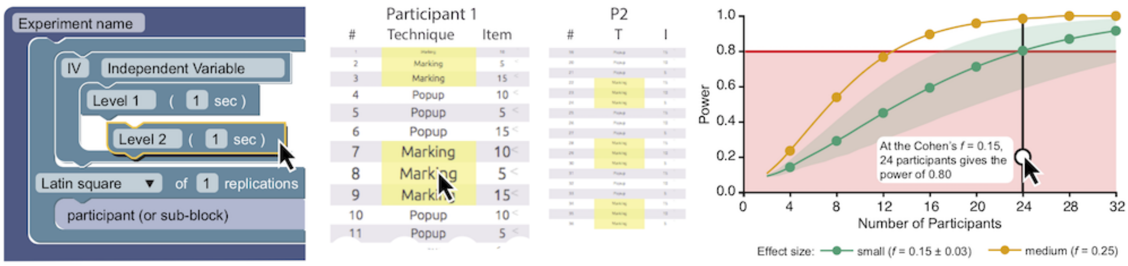


Figure 2. Touchstone2: visual language to specify an experimental design, trial table with fish-eye view, power plot.

paper sticky notes. Story-based interviews reveal both breakdowns and creative new uses of sticky notes. Brainstorming and video brainstorming, informed by the users' stories, generate new ideas. Paper prototyping a design concept related to augmented sticky notes lets designers explore ideas for a future system to address an untapped need or desire. Shooting a video prototype, guided by titlecards and a storyboard, illustrates how future users will interact with the proposed system. Finally, a design walkthrough identifies key problems and suggests ideas for improvement.

At the theoretical level, we have continued our exploration of Information Theory as a design tool for HCI by analyzing past and current applications of Shannon's theory to HCI research to identify areas where information-theoretic concepts can be used to understand, design and optimize human-computer communication [30]. We have also continued our long-standing strand of work on pointing by evaluating several models for assessing pointing performance by participants with motor impairments [27]. Namely, we studied the strengths of weaknesses of various models, from traditional Fitts' Law to the WHo model, the EMG regression and the method of Positional Variance Profiles (PVPs), on datasets from abled participants vs. participants with dyspraxia.

In the context of the ERC ONE project on Unified Principles of Interaction, Philip Tchernavskij defended his Ph.D. thesis on malleable software [40]. The goal of malleable software is to make is as easy as possible for users themselves to change software, or to have it changed on their behalf in response to their developing needs. Current approaches do not address this issue adequately: software engineering promotes flexible code, but in practice this does not help end-users effect change in their software. Based on a study of a network of communities working with biodiversity data, we found that the mode of software production, i.e. the technologies and economic relations that produce software, is biased towards centralized, one-size-fits-all systems. Instead, we should seek to create infrastructures for plurality, i.e. tools that help multiple communities collaborate without forcing them to consolidate around identical interfaces or data representations. Malleable software seeks to maximize the kinds of modifications that can take place through regular interactions, e.g. direct manipulation of interface elements. By generalizing existing control structures for interaction under the concepts of co-occurrences and entanglements, we created an environment where interactions can be dynamically created and modified. The *Tangler* prototype illustrates the power of these concepts to create malleable software.

In collaboration with Aarhus University (Denmark), we created *Videostrates* [22] to explore the notion of an *interactive substrate* for video data. *Videostrates* is based on our joint previous work on *Webstrates* (https://webstrates.net) and supports both live and recorded video composition with a declarative HTML-based notation, combining both simple and sophisticated editing tools that can be used collaboratively. *Videostrates* is programmable and unleashes the power of the modern web platform for video manipulation. We demonstrated its potential through three use scenarios (Fig. 3): collaborative video editing with multiple tools and devices; orchestration of multiple live streams that are recorded and broadcast to a popular streaming platform; and programmatic creation of video using WebGL and shaders for blue screen effects. These scenarios demonstrate *Videostrates*' potential for novel collaborative video editors with fully programmable interfaces.

Figure 3. Videostrates examples: A) Two users collaboratively edit the same videostrate, one with a timeline-based editor and the other with a subtitle editor. The results appear in a live, interactive preview on a large screen. B) Videostrates aggregates, broadcasts and records multiple live streams, here from a statically mounted camera and a smartphone. C) A Videostrate-based computational notebook uses Codestrates to programmatically create a WebGL animation and synchronize its playback with recorded video composited with a green screen.

We conducted an in-depth observational study of landscape architecture students to reveal a new phenomenon in pen-and-touch surface interaction: *interstices* [24]. We observed that bimanual interactions with a pen and touch surface involved various sustained hand gestures, interleaved between their regular commands. Positioning of the non-preferred hand indicates anticipated actions, including: sustained hovering near the surface; pulled back but still floating above the surface; resting in their laps; and stabilizing the preferred hand while handwriting. These intersticial actions reveal anticipated actions and therefore should be taken into account in the design of novel interfaces.

We also started a study of blind or visually impaired people to better understand how they use graphical user interfaces [28]. The goal is to design multimodal interfaces for sighted users that do not rely on the visual channel as much as current GUIs.

In collaboration with the University of Paris Descartes and the ILDA Inria team, we investigated how to help users to query massive data series collections within interaction times. We demonstrated the importance of providing progressive whole-matching similarity search results on large time series collections (100 GB). Our experiments showed that there is a significant gap between the time the 1st Nearest Neighbor (1-NN) is found and the time when the search algorithm terminates [29]. In other words, users often wait without any improvement in their answers. We further showed that high-quality approximate answers are found very early, e.g., in less than one second, so they can support highly interactive visual analysis tasks. We discussed how to estimate probabilistic distance bounds, and how to help analysts evaluate the quality of their progressive results. The results of this collaboration have lead to Gogolou's Ph.D. thesis (ILDA Inria team) [38].

In the context of virtual reality, we explored how to integrate the real world surrounding users in the virtual environment. In many virtual reality systems, user physical workspace is superposed with a particular area of the virtual environment. This spatial consistency allows users to physically walk in the virtual environment and interact with virtual content through tangible objects. However, as soon as users perform virtual navigation to travel on a large scale (i.e. move their physical workspace in the virtual environment), they break this spatial consistency. We introduce two switch techniques to help users to recover the spatial consistency in some predefined virtual areas when using a teleportation technique for the virtual navigation [26]. We conducted a controlled experiment on a box-opening task in a CAVE-like system to evaluate the performance and usability of these switch techniques. The results highlight that helping the user to recover a spatially consistency ensures the accessibility of the entire virtual interaction space of the task. Consequently, the switch techniques decrease time and cognitive effort required to complete the task.

7.2. Human-Computer Partnerships

Participants: Wendy Mackay [correspondant], Baptiste Caramiaux, Téo Sanchez, Carla Griggio, Shu Yuan Hsueh, Wanyu Liu, Joanna Mcgrenere, Midas Nouwens.

ExSitu is interested in designing effective human-computer partnerships, in which expert users control their interaction with technology. Rather than treating the human users as the 'input' to a computer algorithm, we explore human-centered machine learning, where the goal is to use machine learning and other techniques to increase human capabilities. Much of human-computer interaction research focuses on measuring and improving productivity: our specific goal is to create what we call 'co-adaptive systems' that are discoverable, appropriable and expressive for the user.

In creative practices, human-centred machine learning facilitates the workflow for creatives to explore new ideas and possibilities. We compiled recent research and development advances in human-centred machine learning and artificial intelligence (AI), within the field of creative industries, in a white paper commissioned by the NEM (New European Media) initiative [35]. We explored the use of Deep Reinforcement Learning in the context of sound design with sound design experts [37]. We first conducted controlled studies where we compared manual exploration versus exploration by reinforcement. This helped us design a fully working system that we assessed in workshops with expert designers. We showed that an algorithmic sound explorer learning from human preferences enhances the creative process by allowing holistic and embodied exploration as opposed to analytic exploration afforded by standard interfaces.

We also explored how users create their own ecosystems of communication apps as a way to support rich, personalized forms of expression [12]. We wanted to gather data about how people customize apps to enable more personal forms of expression, and how such customizations shape their everyday communication. Given the increasing use of multiple apps with overlapping communication features, we were also interested in how customizing one app influences communication via other apps. We created a taxonomy of customization options based on interviews with 15 “extreme users” of communication apps. We found that participants tailored their apps to express their identities, organizational culture, and intimate bonds with others. They also experienced expression breakdowns: frustrations around barriers to transferring personal forms of expression across apps, which inspired inventive workarounds to maintain cross-app habits of expression, such as briefly switching apps to generate and export content for a particular conversation. We conclude with implications for personalized expression in ecosystems of communication apps.

We investigated the special communication practices between couples [20]. Research shows that sharing streams of contextual information, e.g. location and motion, helps couples coordinate and feel more connected. We studied how couples' communication changes when sharing multiple, persistent information streams. We designed *Lifelines*, a mobile-app technology probe that visualizes up to six streams on a shared timeline: closeness to home, battery level, steps, media playing, texts and calls. A month-long study with nine couples showed that partners interpreted information mostly from individual streams, but also combined them for more nuanced interpretations. Persistent streams allowed missing data to become meaningful and provided new ways of understanding each other. Unexpected patterns from any stream can trigger calls and texts, whereas seeing expected data can replace direct communication, which may improve or disrupt established communication practices.

Finally, we extended our earlier work on the *Expressive Keyboard* by adding animated emojis as a form of expressive output for messaging apps. An initial user study identified both the cumbersome nature of inserting emojis and the creative ways that users construct emoji sequences to convey rich, nuanced non-verbal expressions, including emphasis, change of expressions, and micro stories. We then developed *MojiBoard* [17], an emoji entry technique that lets users generate dynamic parametric emojis from a gesture keyboard. Here, the form of the user's gesture is transformed into an animation, allowing users to “draw” dynamic expressions through their own movements. *MojiBoard* lets users switch seamlessly between typing and parameterizing emojis. *MojiBoard* provides an example of how we can transform a user's gesture into an expressive output, which is reified into an emoji than can be interacted with again.

Wendy Mackay describes how the theoretical foundation of the CREATIV ERC Advance Grant, based on the principle of co-adaptation, influenced her research with musicians, choreographers, graphic designers and other creative professionals. The interview is published in the book “New Directions in Music and Human-Computer Interaction”, Springer Nature, as a chapter entitled “HCI, Music and Art: An Interview with Wendy Mackay” [34]. Along the same lines, she contributed to a chapter “A Design Workbench for Interactive Music Systems” [33] that discusses possible links between the fields of computer music and human-computer interaction (HCI), particularly in the context of the MIDWAY project between Inria, France and McGill University, Canada. The goal of MIDWAY was to construct a “musical interaction design workbench” to facilitate the exploration and development of new interactive technologies for musical creation and performance by bringing together useful models, tools, and recent developments from computer music and HCI. These models and tools have helped expand the possibilities for enhancing musical expression, and provide HCI researchers with a better foundation for the design of tools for “extreme” users.

7.3. Creativity

Participants: Sarah Fdili Alaoui [correspondant], Carla Griggio, Shu Yuan Hsueh, Wendy Mackay, Baptiste Caramiaux, Joanna Mcgrenere, Midas Nouwens, Jean-Philippe Riviere, Nicolas Taffin, Philip Tchernavskij, Theophanis Tsandilas.

ExSitu is interested in understanding the work practices of creative professionals, particularly artists, designers, and scientists, who push the limits of interactive technology. We follow a multi-disciplinary participatory design approach, working with both expert and non-expert users in diverse creative contexts. We also create situations that cause users to reflect deeply on their activities in situ and collaborate to articulate new design problems.

We conducted an interview study of 23 contemporary music composers and choreographers where we focused on the role that physical artifacts play in shaping creative collaborations with performers [13]. We found that creators and performers form relationships where the creator acts as a author, a curator, a planner, or a researcher and the performer acts as an interpreter, a creator, an improvisor, or an informant. Furthermore, we found that creators sculpt, layer and remix artifacts, moving fluidly across these different forms of interaction throughout the creative process.

We studied Kinaesthetic creativity which refers to the body’s ability to generate alternate futures [21]. We probe such creative process by studying how dancers interact with technology to generate ideas. We developed a series of parameterized interactive visuals and asked dance practitioners to use them in generating movement materials. From our study, we define a taxonomy that comprises different relationships and movement responses dancers form with the visuals. We describe resulting types of interaction patterns and demonstrate how dance creativity is driven by the ability to shift between these patterns.

We used technology probes to understand how dancers learned dance fragments from videos [15]. We introduced *MoveOn*, which lets dancers decompose video into short, repeatable clips to support their learning. This served as an effective analysis tool for identifying the changes in focus and understanding dancers decomposition and recomposition processes. Additionally we compared the teacher’s and dancers’ decomposition strategies, and how dancers learn on their own compared to teacher-created decompositions. We found that they all ungroup and regroup dance fragments, but with different foci of attention, which suggests that teacher-imposed decomposition is more effective for introductory dance students, whereas personal decomposition is more suitable for expert dancers.

We ran a workshop [25] at ACM *Creativity and Cognition* that explored how distributed forms of creativity arising in play can help guide and foster supportive research, game design, and technology. We brought together researchers, game designers, and others to examine theories of creativity and play, game design practices, and methods for studying creativity.

We developed a taxonomy [18] on technologies using Defamiliarization to to support Co-Creation in choreographic practices. Regarding intersection of choreographic practice and HCI, Sarah Fdili Alaoui [16] describe her research and creation journey of an interactive choreographic dance piece called SKIN. This generated a set of research questions that she addresses through experience explicitation interviews of both audience and creative team members on the lived experience of making and attending the performance and the emergent relationships between dance, media and interaction as well as the tensions and negotiations that emerged from integrating technology in art. She discusses her approach as anti-solutionist and argue for more openness in HCI to allow artists to contribute.

Finally, we assessed the inter-rater reliability of the Laban Movement Analysis system used in choreography and dance notation [11].

7.4. Collaboration

Participants: Cédric Fleury [correspondant], Michel Beaudouin-Lafon, Wendy Mackay, Carla Griggio, Yujiro Okuya, Arthur Fages.

ExSitu explores new ways of supporting collaborative interaction and remote communication. In particular, we studied co-located collaboration on large wall-sized display, video-conferencing systems for remote collaboration, and collaboration between professional designers and developers during the design of interactive systems.

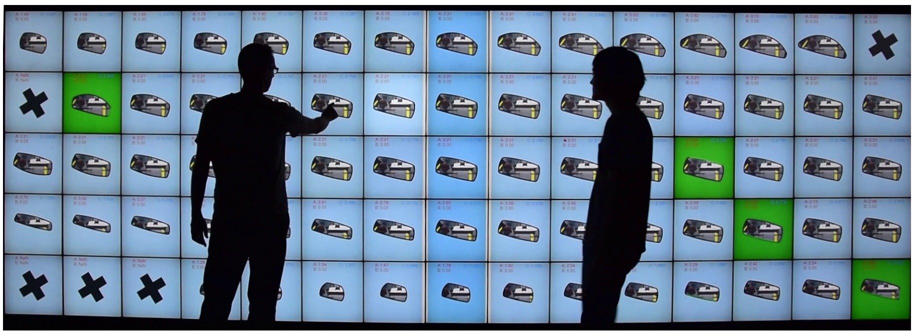


Figure 4. Collaborative exploration of multiple design alternatives of a car rear-view mirror on a wall-sized display.

Multi-touch wall-sized displays, as those of the Digiscope network (http://digiscope.fr/, afford collaborative exploration of large datasets and re-organization of digital content. In the context of industrial design, computer-aided design (CAD) is now an essential part of the design process allowing experts to evaluate and adjust product design using digital mock-ups. We investigated how a wall-sized display could be used to allow multidisciplinary collaborators (e.g. designers, engineers, ergonomists) to explore large number of design alternatives. In particular, we design a system which allows non-CAD expert to generate and distribute on a wall-sized display multiple various of a CAD model (Figure 4). We ran a usability study and a controlled experiment to assess the benefit of wall-sized displays in such context. Yujiro Okuya, under the supervision of Patrick Bourdot (LIMSI-CNRS) and Cédric Fleury, successfully defended his thesis *CAD Modification Techniques for Design Reviews on Heterogeneous Interactive Systems* [39] on this topic.

For remote collaboration using video, interpreting gaze direction is critical for communication between coworkers sitting around a table and a remote satellite colleague. However, 2D video distorts images and makes this interpretation inaccurate. We proposed GazeLens [23], a video conferencing system that improves coworkers' ability to interpret the satellite worker's gaze (Figure 5). A 360 camera captures the coworkers and a ceiling camera captures artifacts on the table. The system combines these two video feeds in an interface. Lens widgets strategically guide the satellite worker's attention toward specific areas of her/his screen allowing coworkers to clearly interpret her/his gaze direction. Controlled experiments showed that GazeLens increases coworkers' overall gaze interpretation accuracy in comparison to a conventional video conferencing system.

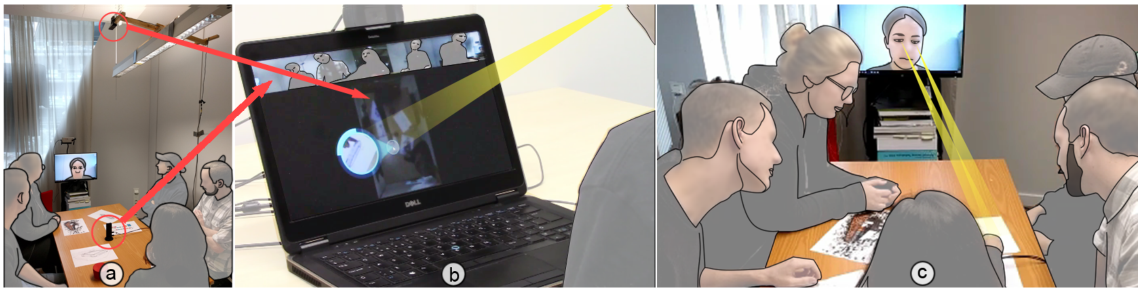


Figure 5. GazeLens system. (left) On the coworkers' side, a 360 camera on the table captures coworkers and a webcam mounted on the ceiling captures artifacts on the table. (middle) Video feeds from the two cameras are displayed on the screen of the remote satellite worker; a virtual lens strategically guides her/his attention towards a specific screen area according to the observed artifact. (right) The satellite's gaze, guided by the virtual lens, is aligned towards the observed artifact on the coworkers' space.

Finally, we also conducted an in-depth study of the collaboration patterns between designers and developers of interactive systems, and created a tool, *Enact*, to facilitate their work [14]. Professional designers and developers often struggle when transitioning between the design and implementation of an interactive system. We found that current practices induce unnecessary rework and cause discrepancies between design and implementation. We identified three recurring types of breakdowns: omitting critical details, ignoring edge cases, and disregarding technical limitations. We introduced four design principles to create tools that mitigate these problems: Provide multiple viewpoints, maintain a single source of truth, reveal the invisible and support design by enaction. We applied these principles to create *Enact*, a live environment for prototyping touch-based interactions (Fig. 6). We conducted two studies to assess *Enact* and compare it with current tools. Results suggest that *Enact* helps participants detect more edge cases, increases designers' participation and provides new opportunities for co-creation.

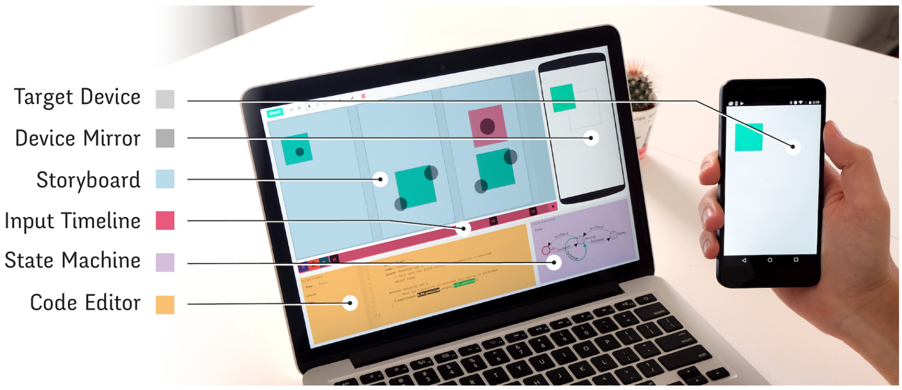


Figure 6. Enact uses a target mobile device and a desktop interface with five areas: a storyboard with consecutive screens, an event timeline with a handle for each screen, a state machine, a code editor and a device mirror.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Virtual Reality for Interacting with Building Information Model at Paris-Saclay

Type: Equipment and human resources

Funding: STIC Paris-Saclay

Duration: 2018-2019

Coordinator: Jean-Marc Vézien (LIMSI-CNRS)

Partners: CNRS, Univ. Paris-Sud

Inria contact: Cédric Fleury

Abstract: The goal of this project is to develop interactive tools for BIM application in virtual reality using a user-centered design approach. The project will use as a case study the interior design of the *Learning Center* building on Paris-Saclay campus.

8.1.2. Projet numérique du Learning Center de l'Université

Type: Equipment and subcontracting

Funding: Learning Center Paris-Saclay

Duration: 2019

Coordinator: Michel Beaudouin-Lafon

Partners: Univ. Paris-Sud

Inria contact: Michel Beaudouin-Lafon

Abstract: The goal of this project (30k) is to create an interactive installation presenting the portraits of Ph.D. students from Université Paris-Saclay. It is a collaboration with portrait photographer Didier Goupy. The installation is designed to be exhibited in various sites of Université Paris-Saclay until it is permanently installed in the Learning Center of Université Paris-Saclay. The project was presented at the Ph.D. graduation ceremony of Université Paris-Saclay in June, 2019, and at the Fête de la Science in October, 2019, and will be permanently exhibited in the future Learning Center of Université Paris-Saclay.

8.1.3. Living Archive

Type: Equipment and human resources

Funding: STIC department grant

Duration: 2019-2020

Coordinator: Sarah Fdili Alaoui

Partners: Learning Center

Inria contact: Sarah Fdili Alaoui

Abstract: The project's ambition is to design interactive systems that allow practioners to easily document their dance using their own methods and personal artifacts emphasizing a first-person perspective and minimizing imposed choices from academic researchers.

8.2. National Initiatives

8.2.1. ANR

ELEMENT: Enabling Learnability in Human Movement Interaction

Type: Equipment and human resources

Funding: ANR

Duration: 2019-2022

Coordinator: Baptiste Caramiaux, Sarah Fdili Alaoui, Wendy Mackay

Partners: IRCAM, LIMSI

Inria contact: Baptiste Caramiaux

Abstract: The goal of this project is to foster innovation in multimodal interaction, from non-verbal communication to interaction with digital media/content in creative applications, specifically by addressing two critical issues: the design of learnable gestures and movements; and the development of interaction models that adapt to a variety of user's expertise and facilitate human sensorimotor learning.

8.2.2. Investissements d'Avenir

8.2.2.1. Digiscope - Collaborative Interaction with Complex Data and Computation

Type: EQUIPEX (Equipement d'Excellence)

Duration: 2011-2019

Coordinator: Michel Beaudouin-Lafon

Partners: Université Paris-Saclay (coordinator), Université Paris-Sud, CNRS, CEA, Inria, Institut Mines-Telecom, CentraleSupelec, Université Versailles - Saint-Quentin, ENS Paris-Saclay, Maison de la Simulation

Overall budget: 22.5 Meuros, including 6.7 Meuros public funding from ANR

Abstract: The goal of the project is to create ten high-end interactive rooms interconnected by high-speed networks and audio-video facilities to support remote collaboration across interactive visualization environments. The equipment will be open to outside users and targets four main application areas: scientific discovery, product lifetime management, decision support for crisis management, and education and training. Digiscope includes the existing WILD room, and funded the WILDER room. ExSitu contributes its expertise in the design and evaluation of advanced interaction techniques and the development of distributed software architectures for interactive systems. All ten rooms and the telepresence network are operational. The project was successfully evaluated by an international jury in June, 2017.

8.3. European Initiatives

8.3.1. European Research Council (ERC)

8.3.1.1. Creating Human-Computer Partnerships

Program: ERC Advanced Grant

Project acronym: CREATIV

Project title: Creating Human-Computer Partnerships

Duration: June 2013 - May 2019

Coordinator: Wendy Mackay

Abstract: CREATIV explores how the concept of co-adaptation can revolutionize the design and use of interactive software. Co-adaptation is the parallel phenomenon in which users both adapt their behavior to the system's constraints, learning its power and idiosyncrasies, and appropriate the system for their own needs, often using it in ways unintended by the system designer. A key insight in designing for co-adaptation is that we can encapsulate interactions and treat them as first class objects, called interaction instruments This lets us focus on the specific characteristics of how human users express their intentions, both learning from and controlling the system. By making instruments co-adaptive, we can radically change how people use interactive systems, providing incrementally learnable paths that offer users greater expressive power and mastery of their technology. The initial goal of the CREATIV project is to fundamentally improve the learning and expressive capabilities of advanced users of creative software, offering significantly enhanced methods for expressing and exploring their ideas. The ultimate goal is to radically transform interactive systems for everyone by creating a powerful and flexible partnership between human users and interactive technology.

8.3.1.2. Unified Principles of Interaction

Program: ERC Advanced Grant

Project acronym: ONE

Project title: Unified Principles of Interaction

Duration: October 2016 - September 2020

Coordinator: Michel Beaudouin-Lafon

Abstract: The goal of ONE is to fundamentally re-think the basic principles and conceptual model of interactive systems to empower users by letting them appropriate their digital environment. The project addresses this challenge through three interleaved strands: empirical studies to better understand interaction in both the physical and digital worlds, theoretical work to create a conceptual model of interaction and interactive systems, and prototype development to test these principles and concepts in the lab and in the field. Drawing inspiration from physics, biology and psychology, the conceptual model combines *substrates* to manage digital information at various levels of abstraction and representation, *instruments* to manipulate substrates, and *environments* to organize substrates and instruments into digital workspaces.

8.3.1.3. Humane AI (801)

Title: Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us

Program: FET Flagships

Duration: March 2019 - February 2020

Coordinator: DFKI (Germany)

Partners:

- Aalto Korkeakoulusaatio SR (Finland)
- Agencia Estatal Consejo Superior De Investigaciones Cientificas (Spain)
- Albert-ludwigs-universitaet Freiburg (Germany)
- Athina-erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis (Greece)
- Consiglio Nazionale Delle Ricerche (Italy)
- Deutsches Forschungszentrum Fur Kunstliche Intelligenz GMBH (Germany)
- Eidgenoessische Technische Hochschule Zürich (Switzerland)
- Fondazione Bruno Kessler (Italy)
- German Entrepreneurship GMBH (Germany)
- INESC TEC - Instituto De Engenharia De Sistemas E Computadores, Tecnologia E Ciencia (Portugal)
- ING Groep NV (Netherlands)
- Institut Jozef Stefan (Slovenia)
- Institut Polytechnique De Grenoble (France)
- Knowledge 4 All Foundation LBG (United Kingdom)
- Kobenhavns Universitet (Denmark)
- Kozep-europai Egyetem (Hungary)
- Ludwig-maximilians-universitaet Muenchen (Germany)
- Max-planck-gesellschaft Zur Forderung Der Wissenschaften EV (Germany)
- Technische Universitaet Kaiserslautern (Germany)
- Technische Universitaet Wien (Austria)

Technische Universitat Berlin (Germany)

Technische Universiteit Delft (Netherlands)

Thales SIX GTS FRANCE SAS (France)

The University Of Sussex (United Kingdom)

Universidad Pompeu Fabra (Spain)

Universita Di Pisa (Italy)

Universiteit Leiden (Netherlands)

University College Cork - National University Of Ireland, Cork (Ireland)

Uniwersytet Warszawski (Poland)

Volkswagen AG (Germany)

Inria contact: Wendy Mackay

The presence and capabilities of artificial intelligence (AI) have grown significantly and will continue to do so. The Humane AI Flagship will develop the scientific foundations and technological breakthroughs needed to shape the ongoing AI revolution. The goal is to deploy AI systems that enhance human capabilities and empower individuals and societies, and ultimately extend human intelligence (rather than replace it). With 35 partners from 17 countries, Humane AI is undertaking a preparatory action to draft an ambitious research agenda to provide competitive advantages to European industry and substantial benefits to society. Partners are united by the vision of a new generation of ethical, value-oriented, and human-centric European approach to AI.

8.4. International Initiatives

8.4.1. Participation in Other International Programs

8.4.1.1. Inria International Chairs

IIC MCGRENERE Joanna

Title: Personalization through Co-Adaptive Human-Computer Interaction

International Partner (Institution - Laboratory - Researcher):

University of British Columbia (Canada) - Dept of Computer Science - Joanna McGrenere

Duration: 2017 - 2021

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Joanne McGrenere, Professor at the University of British Columbia, Canada and Inria Chair, visited for two months, to work with Wendy Mackay and Michel Beaudouin-Lafon.

Susanne Bødker, Professor at Aarhus University, Denmark, visited for a week to work with Wendy Mackay and Michel Beaudouin-Lafon.

8.5.1.1. Internships

Injung Lee, Ph.D. student from KAIST, South Korea, visited for five months to work with Michel Beaudouin-Lafon.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

9.1.1.1. Member of the Conference Program Committees

- ACM CHI 2020, *ACM CHI Conference on Human Factors in Computing Systems*: Theophanis Tsandilas
- ACM UIST 2019, *ACM Symposium on User Interface Software and Technology*: Michel Beaudouin-Lafon, Wendy Mackay
- ACM VRST 2019, *ACM Symposium on Virtual Reality Software and Technology*: Cédric Fleury
- EuroVR 2019, *EuroVR International Conference*: Cédric Fleury
- MOCO 2019, *International Conference on Movement and Computing*: Sarah Fdili Alaoui, Baptiste Caramiaux
- IHM 2019, *Conférence Francophone d'Interaction Homme-Machine*: Sarah Fdili Alaoui

9.1.1.2. Reviewer

- ACM CHI 2019-20, *ACM CHI Conference on Human Factors in Computing Systems*: Sarah Fdili Alaoui, Cédric Fleury, Baptiste Caramiaux, Michel Beaudouin-Lafon, Wendy Mackay, Julien Gori, Alexander Eiselmayer
- ACM UIST 2019, *ACM Symposium on User Interface Software and Technology*: Theophanis Tsandilas, Cédric Fleury, Julien Gori
- ACM DIS 2019, *Designing Interactive Systems*: Baptiste Caramiaux, Sarah Fdili Alaoui
- ACM ACII, : Baptiste Caramiaux
- IEEE VIS 2019, *IEEE Visualization Conference*: Theophanis Tsandilas, Michel Beaudouin-Lafon
- IEEE VR 2019-20, *Virtual Reality Conference*: Cédric Fleury
- ACM ISS 2019, *ACM International Conference on Interactive Surfaces and Spaces*: Theophanis Tsandilas
- ACM MobileHCI (2019) *The Conference on Human-Computer Interaction with Mobile Devices and Services*: Julien Gori
- ACM NordiCHI (2018) *The Nordic Conference on Human-Computer Interaction*: Julien Gori
- EURASP Eusipco (2019) *European Signal Processing Conference*: Julien Gori
- IEEE SMC (2018) *IEEE International Conference on Systems, Man, and Cybernetics*: Julien Gori
- ACM CHIPLAY 2019, *ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play*: Viktor Gustafsson
- IHM 2019, *Conférence Francophone d'Interaction Homme-Machine*: Theophanis Tsandilas, Wendy Mackay

9.1.2. Journal

9.1.2.1. Member of Editorial Boards

- Editor for the Human-Computer Interaction area of the ACM Books Series (published with Morgan & Claypool Publishers): Michel Beaudouin-Lafon (2013-)
- TOCHI, *Transactions on Computer Human Interaction*, ACM: Michel Beaudouin-Lafon (2009-), Wendy Mackay (2016-), Baptiste Caramiaux (2019-)
- PloS ONE: Baptiste Caramiaux (2018-)
- JIPS, *Journal d'Interaction Personne-Système*, AFIHM: Michel Beaudouin-Lafon (2009-)

- Frontiers in *Virtual Reality*: Cédric Fleury (2019-)

9.1.2.2. Reviewer - Reviewing Activities

- TOCHI, *Transactions on Computer Human Interaction*, ACM: Theophanis Tsandilas, Julien Gori
- ACM Books Series: Theophanis Tsandilas
- Frontiers in *Robotics and AI*, Virtual Environments section: Cédric Fleury

9.1.3. Invited Talks

- CIO Workshop, Sanderberg, Denmark *Generative Theories of Interaction (in the making)*, 23 January 2019: Michel Beaudouin-Lafon, Susanne Bødker, and Wendy Mackay
- Colloque CNRS Humain et Numérique en Interaction, Paris, *Aura-t-on encore besoin d'interagir avec des ordinateurs ?*, February 2019: Michel Beaudouin-Lafon
- Stanford University, Palo Alto CA *Human-Computer Partnerships*, 27 February 2019: Wendy Mackay
- IHM & AI Workshop, Paris *Partenariats Homme-Machines*, 14 March 2019: Wendy Mackay
- DYSTOPIA, Paris, *L'IA nous rendra-t-elle inhumains ?*, March 2019: Michel Beaudouin-Lafon
- Stanford University, *Unified Principles of Interaction*, March 2019: Michel Beaudouin-Lafon
- CHI 2019 Panel, Glasgow, Scotland *Rigor, Relevance and Impact: The Tensions and Trade-Offs Between Research in the Lab and in the Wild*, April 2019: Wendy Mackay
- MIT CSAIL, *Unified Principles of Interaction*, May 2019: Michel Beaudouin-Lafon
- SystemX, Paris *Partenariats Homme-Machines*, 29 May 2019: Wendy Mackay
- ESPGG, Open Lab Days, *Human-centred Machine Learning in Creative Applications*, June 2019: Baptiste Caramiaux
- University of London Institute, *AI in the Media and Creative Industries*, June 2019: Baptiste Caramiaux
- Aachen University Graduation Keynote *Human-Computer Partnerships*, 17 June 2019: Wendy Mackay
- Creative Industries & AI, Paris *Human-Computer Partnerships*, 21 June 2019: Wendy Mackay
- User Interfaces group, Aalto University, Finland, *A new take on the speed-accuracy tradeoff of Aimed Movements – Implications and Applications for HCI*, October 2019: Julien Gori
- ACM UIST Symposium, New Orleans, Visions talk *A World Without Apps*, October 2019: Michel Beaudouin-Lafon
- AI & Gender Equality Seminar, UNESCO, Paris *Does AI introduce new risks and opportunities for gender equality?*, 2 November, 2019: Wendy Mackay
- LIMSI, *Human-centred Machine Learning*, Novembre 2019: Baptiste Caramiaux
- University of Toronto TUX Sanders Series Talk, *Unified Principles of Interaction*, November 2019: Michel Beaudouin-Lafon
- IRCAM Paris, *State of The art of tools supporting dance learning*, ELEMENT workshop, 2019: Sarah Fdili Alaoui
- ENSAD, Paris, *Live Coding with interactive movement based technologies*, Virtual creativity collective realities, 2019: Sarah Fdili Alaoui
- Aarhus, Danemark, *Live Coding with interactive movement based technologies*, Performing Art Platform, 2019: Sarah Fdili Alaoui
- Aarhus University, Danemark, *Crafting dance and technologies*, 2019: Sarah Fdili Alaoui
- Conservatory of Vicenza, Italy, Workshop on *Music dance and interaction*, 2019: Sarah Fdili Alaoui
- Conservatory of Vicenza, Italy, Master class of *Creation with technologies*, 2019: Sarah Fdili Alaoui

- Stereolux, Nantes, France, Workshop on *Live Coding with interactive movement based technologies* at LAB-DAYS - MovA, 2019: Sarah Fdili Alaoui

9.1.4. Leadership within the Scientific Community

- Information Science and Technology (STIC) Department, Université Paris-Saclay: Michel Beauouin-Lafon (chair since June 2018), Wendy Mackay (member)
- Research division, Université Paris-Saclay: Michel Beaudouin-Lafon (advisor for Digital Sciences since June 2018)
- Digiteo RTRA research network, Université Paris-Saclay: Michel Beauouin-Lafon (director since June 2018)
- Computer Science Department, Université Paris-Sud: Michel Beaudouin-Lafon (vice-President for research)
- ACM Technology Policy Council: Michel Beaudouin-Lafon (vice-chair)

9.1.5. Scientific Expertise

International

- HCERES Evaluation: Wendy MAckay, external expert
- NSERC Evaluation: Michel Beaudouin-Lafon, external expert
- ACM SIGCHI "Lifetime Service Award" committee chair: Michel Beaudouin-Lafon
- ACM "Policy Award" committee adjunct chair: Michel Beaudouin-Lafon

National

- CNRS INS2I "Cellule ERC": Michel Beaudouin-Lafon, member
- Agence Nationale de la Recherche (ANR), Appel à projets génériques: Sarah Fdili Alaoui, reviewer

9.1.6. Research Administration

Telecom ParisTech, "Comité de la recherche": Michel Beaudouin-Lafon (member)

"Conseil de Laboratoire", LRI: Wendy Mackay, Cédric Fleury (members)

"Commission consultatives paritaires (CCP)" Inria: Wendy Mackay (President)

"Conseil Scientifique", LRI: Michel Beaudouin-Lafon (member)

CCSU, "Commission Consultative de Spécialistes de l'Université", Université Paris-Sud: Michel Beaudouin-Lafon, Wendy Mackay (members)

"Commission Locaux", LRI: Theophanis Tsandilas (member)

"Commission Scientifique", Inria: Theophanis Tsandilas (member)

"Jury de recrutement, Maître de Conférences", Télécom ParisTech: Michel Beaudouin-Lafon (member)

"Assessment committee for associate professors", Aarhus University: Michel Beaudouin-Lafon (member)

"Comité de sélection, Maître de Conférences", Université Paris-Sud: Cédric Fleury

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

International Masters: Theophanis Tsandilas, *Probabilities and Statistics*, 32h, M1, Univ. Paris-Saclay

Interaction & HCID Masters & Innovation and Entrepreneurship Minor: Sarah Fdili Alaoui, Coordinator

HCID Masters: Sarah Fdili Alaoui, *Innovation & Entrepreneurship Advanced*, 15h, M2, Univ. Paris-Sud

HCID Masters: Michel Beaudouin-Lafon, Wendy Mackay, *Fundamentals of Situated Interaction*, 21 hrs, M1/M2, Univ. Paris-Sud

Interaction & HCID Masters: Sarah Fdili Alaoui, *Creative Design*, 27h, M1 et M2, Univ. Paris-Sud

Interaction & HCID Masters: Michel Beaudouin-Lafon, *Fundamentals of Human-Computer Interaction*, 21 hrs, M1/M2, Univ. Paris-Sud

Interaction & HCID Masters: Michel Beaudouin-Lafon & Cédric Fleury, *Groupware and Collaborative Interaction*, 31.5 hrs, M1/M2, Univ. Paris-Sud

Interaction & HCID Masters: Wendy Mackay, *Career Seminar* 6 hrs, M2, Univ. Paris-Sud

Interaction & HCID Masters: Wendy Mackay, *Design of Interactive Systems*, 21 hrs, M1/M2, Univ. Paris-Sud

Interaction & HCID Masters: Wendy Mackay, *Advanced Design of Interactive Systems*, 21 hrs, M1/M2, Univ. Paris-Sud

Interaction & HCID Masters: Baptiste Caramiaux, *Gestural and Mobile Interaction*, 24 hrs, M1/M2, Univ. Paris-Sud

Polytech: Cédric Fleury, *Projet Java-Graphique-IHM*, 24 hrs, 3st year, Univ. Paris-Sud

Polytech: Cédric Fleury, *Interaction Homme-Machine*, 18 hrs, 3st year, Univ. Paris-Sud

Polytech: Cédric Fleury, *Option Réalité Virtuelle*, 56 hrs, 5th year, Univ. Paris-Sud

Polytech: Cédric Fleury, *Réalité Virtuelle et Interaction*, 48 hrs, “Apprentis” 5th year, Univ. Paris-Sud

9.2.2. Supervision

PhD students

PhD: Philip Tchernavskij, *Designing and Programming Malleable Software*, Université Paris-Saclay, 3 December 2019. Advisors: Michel Beaudouin-Lafon

PhD: Yujiro Okuya, *CAD Modification Techniques for Design Reviews on Heterogeneous Interactive Systems*, 8 November 2019. Advisors: Patrick Bourdot (LIMSI-CNRS) & Cédric Fleury

PhD: Anna Gogolou, *Iterative and Expressive Querying for Big Data Series*, 15 November 2019. Advisors: Anastasia Bezerianos (ILDA-Inria), Themis Palpanas (Université Paris-Descartes) & Theophanis Tsandilas

PhD in progress: Stacy (Shu-Yuan) Hsueh, *Embodied design for Human-Computer Co-creation*, November 2015. Advisors: Wendy Mackay & Sarah Fdili Alaoui

PhD in progress: Jean-Philippe Rivière, *Embodied Design for Human-Computer Partnership in Learning Contexts*, October 2017. Advisors: Wendy Mackay, Sarah Fdili Alaoui & Baptiste Caramiaux

PhD in progress: Yiran Zhang, *Telepresence for remote and heterogeneous Collaborative Virtual Environments*, October 2017. Advisors: Patrick Bourdot (LIMSI-CNRS) & Cédric Fleury

PhD in progress: Téo Sanchez, *Co-Learning in Interactive Systems*, September 2018. Advisors: Baptiste Caramiaux & Wendy Mackay

PhD in progress: Elizabeth Walton, *Inclusive Design in Embodied Interaction*, November 2018. Advisor: Wendy Mackay

PhD in progress: Miguel Renom, *Theoretical bases of human tool use in digital environments*, October 2018. Advisors: Michel Beaudouin-Lafon & Baptiste Caramiaux

PhD in progress: Han Han, *Participatory design of digital environments based on interaction substrates*, October 2018. Advisor: Michel Beaudouin-Lafon

PhD in progress: Viktor Gustafsson, *Co-adaptive Instruments fo Game Design*, October 2018. Advisor: Wendy Mackay

PhD in progress: Yi Zhang, *Generative Design using Instrumental Interaction, Substrates and Co-adaptive Systems*, October 2018. Advisor: Wendy Mackay

PhD in progress: Martin Tricaud, *Instruments and Substrates for Procedural Creation Tools*, October 2019. Advisor: Michel Beaudouin-Lafon

PhD in progress: Arthur Fages, *Collaborative 3D Modeling in Augmented Reality Environments*, December 2019. Advisors: Cédric Fleury & Theophanis Tsandilas

Masters students

Clément Sauvard, "*Designing 3D Scenes with multiple AR views*": Cédric Fleury & Theophanis Tsandilas

Cyril Crebouw, "*Remote collaboration across wall-sized displays*": Cédric Fleury & Michel Beaudouin-Lafon

Ellen Sigloch, "*Prototips website*": Wendy Mackay

Yuyan Jing, "*Coadaptive Map: A Design tool for GIS users*": Wendy Mackay

Alina Nicolae, "*TweakyTemplate: Reifying template tweaks for book design*": Wendy Mackay

Wuji Geng, "*Studies of the Interaction Museum: The gap between HCI research and practice*": Wendy Mackay

Xi Hu, "*Motion Palette: Motion Exploration Support Tool for Dynamic Visual Effects*": Wendy Mackay

Siba Siddique, "*Investigating User Decision in Human-AI Interaction*": Baptiste Caramiaux

9.2.3. Juries

PhD theses

Sylvain Pauchet, University of Toulouse/ENAC "*From surface to surface: Transformations de surface tactile pour l'interaction incarnée dans le cockpit*" (advisor: Stéphane Conversy): Wendy Mackay, reviewer

Hugo Scurto, IRCAM/Sorbonne University "*Designing With Machine Learning for Interactive Music Dispositifs*" (advisor: Frédéric Bevilacqua): Wendy Mackay, reviewer

Emmanouil Giannisakis, Télécom ParisTech, "*Promoting and characterizing the menu to keyboard shortcuts transition*" (advisor: Gilles Bailly): Theophanis Tsandilas, examiner

Julien Casarin, Université de Strasbourg, "*Proposition d'un protocole web pour la collaboration multi-support en environnement 3D*" (advisor: Dominique Bechman): Michel Beaudouin-Lafon, president

Emmanouil Potetsianakis, Télécom ParisTech, "*Amélioration des applications vidéo grâce aux métadonnées temporelles*" (advisor: Jean Le Feuvre): Michel Beaudouin-Lafon, examiner

Maxime Garcia, Université Grenoble Alpes, Inria Rhones Alpes, "*Transfert d'animation : animer des personnages virtuels par le jeu et le mime*" (advisor: Rémi Ronfard): Sarah Fdili Alaoui, invited examiner

9.3. Popularization

9.3.1. Articles and contents

- Radio show "Les P'tits Bateaux" on France Inter public radio: Michel Beaudouin-Lafon (interviews in response to children questions on 5 May, 30 June, 18 August, 24 November 2019)

9.3.2. Education

- Chapter about the Web in a textbook for high-school students: Michel Beaudouin-Lafon [36]

9.3.3. Interventions

- Presentation of research work on dance movement during the dance show “Frame(d)”, at Micadance (25 May 2019) and SUAPS Paris-Sud (June 2019): Rivière Jean-Philippe
- Université Paris-Saclay Learning Center: art-science project with photographer Didier Goupy presenting the portraits of Ph.D students of Université Paris-Saclay. Ph.D. graduation ceremony, Université Paris-Saclay, June 2019; Fête de la Science, Gif-sur-Yvette, October 2019

10. Bibliography

Major publications by the team in recent years

[1] S. F. ALAOUI, J. FRANÇOISE, T. SCHIPHORST, K. STUDD, F. BEVILACQUA.*Seeing, Sensing and Recognizing Laban Movement Qualities*, in "ACM Conference on Human Factors in Computing Systems (CHI)", Denver, United States, May 2017 [*DOI :* 10.1145/3025453.3025530], https://hal.inria.fr/hal-01663132

[2] J. ALVINA, J. MALLOCH, W. MACKAY.*Expressive Keyboards: Enriching Gesture-Typing on Mobile Devices*, in "Proceedings of the 29th ACM Symposium on User Interface Software and Technology (UIST 2016)", Tokyo, Japan, ACM, October 2016, p. 583 - 593 [*DOI :* 10.1145/2984511.2984560], https://hal.inria.fr/hal-01437054

[3] I. AVELLINO, C. FLEURY, W. MACKAY, M. BEAUDOUIN-LAFON.*CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays*, in "Proceedings of the CHI Conference on Human Factors in Computing Systems", Denver, United States, CHI ’17, ACM, May 2017, p. 6718 - 6729 [*DOI :* 10.1145/3025453.3025604], https://hal.archives-ouvertes.fr/hal-01544645

[4] B. CARAMIAUX, F. BEVILACQUA, M. M. WANDERLEY, C. PALMER.*Dissociable effects of practice variability on learning motor and timing skills*, in "PLoS ONE", 2018, vol. 13, n^o 3, e0193580 [*DOI :* 10.1371/JOURNAL.PONE.0193580], https://hal.archives-ouvertes.fr/hal-01726186

[5] M. C. CIOLFI FELICE, N. MAUDET, W. MACKAY, M. BEAUDOUIN-LAFON.*Beyond Snapping: Persistent, Tweakable Alignment and Distribution with StickyLines*, in "UIST ’16 Proceedings of the 29th Annual Symposium on User Interface Software and Technology", Tokyo, Japan, Proceedings of the 29th Annual Symposium on User Interface Software and Technology, October 2016 [*DOI :* 10.1145/2984511.2984577], https://hal.archives-ouvertes.fr/hal-01410171

[6] C. KLOKMOSE, J. EAGAN, S. BAADER, W. MACKAY, M. BEAUDOUIN-LAFON.*Webstrates: Shareable Dynamic Media*, in "ACM Symposium on User Interface Software and Technology (UIST)", Charlotte, United States, UIST ’15 Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, ACM, November 2015, p. 280-290 [*DOI :* 10.1145/2807442.2807446], https://hal.archives-ouvertes.fr/hal-01242672

[7] W. LIU, R. LUCAS D’OLIVEIRA, M. BEAUDOUIN-LAFON, O. RIOUL.*BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation*, in "ACM CHI 2017 - International conference of Human-Computer Interaction", Denver, United States, May 2017, p. 5869-5880 [*DOI :* 10.1145/3025453.3025524], https://hal.inria.fr/hal-01677122

[8] Y. OKUYA, N. LADEVEZE, C. FLEURY, P. BOURDOT.*ShapeGuide: Shape-Based 3D Interaction for Parameter Modification of Native CAD Data*, in "Frontiers in Robotics and AI", November 2018, vol. 5 [*DOI :* 10.3389/FROBT.2018.00118], https://hal.inria.fr/hal-01958226

[9] T. TSANDILAS.*Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation*, in "ACM Transactions on Computer-Human Interaction", June 2018, vol. 25, n^o 3, p. 1-49 [*DOI :* 10.1145/3182168], https://hal.archives-ouvertes.fr/hal-01788775

[10] M. WESSELY, T. TSANDILAS, W. MACKAY.*Stretchis: Fabricating Highly Stretchable User Interfaces*, in "ACM Symposium on User Interface Software and Technology (UIST)", Tokyo, Japan, October 2016, p. 697-704 [*DOI :* 10.1145/2984511.2984521], https://hal.archives-ouvertes.fr/hal-01412289

Publications of the year

Articles in International Peer-Reviewed Journal

[11] U. BERNARDET, S. F. ALAOUI, K. STUDD, K. BRADLEY, P. PASQUIER, T. SCHIPHORST.*Assessing the reliability of the Laban Movement Analysis system*, in "PLoS ONE", June 2019 [*DOI :* 10.1371/JOURNAL.PONE.0218179], https://hal.inria.fr/hal-02422169

[12] C. GRIGGIO, J. MCGRENERE, W. MACKAY.*Customizations and Expression Breakdowns in Ecosystems of Communication Apps*, in "Proceedings of the ACM on Human-Computer Interaction ", November 2019, vol. 3, n^o CSCW, 26 [*DOI :* 10.1145/3359128], https://hal.archives-ouvertes.fr/hal-02295535

[13] S. HSUEH, S. FDILI ALAOUI, W. MACKAY.*Deconstructing Creativity: Non-Linear Processes and Fluid Roles in Contemporary Music and Dance*, in "Proceedings of the ACM on Human-Computer Interaction ", November 2019, vol. 3, n^o 203, 21 [*DOI :* 10.1145/3359305], https://hal.archives-ouvertes.fr/hal-02430819

[14] G. LEIVA, N. MAUDET, W. MACKAY, M. BEAUDOUIN-LAFON.*Enact: Reducing Designer-Developer Breakdowns when Prototyping Custom Interactions*, in "ACM Transactions on Computer-Human Interaction", June 2019, vol. 26, n^o 3, p. 1-48 [*DOI :* 10.1145/3310276], https://hal.archives-ouvertes.fr/hal-02218117

[15] J.-P. RIVIÈRE, S. F. ALAOUI, B. CARAMIAUX, W. MACKAY.*Capturing Movement Decomposition to Support Learning and Teaching in Contemporary Dance*, in "Proceedings of the ACM on Human-Computer Interaction ", August 2019, vol. 3, n^o CSCW, p. 1-22 [*DOI :* 10.1145/3359188], https://hal.archives-ouvertes.fr/hal-02378487

International Conferences with Proceedings

[16] S. F. ALAOUI.*Making an Interactive Dance Piece: Tensions in Integrating Technology in Art*, in "Designing Interactive Systems Conference (DIS'19)", San Diego, United States, June 2019, p. 1195–1208, https://hal.inria.fr/hal-02434441

[17] J. ALVINA, C. QU, J. MCGRENERE, W. MACKAY.*MojiBoard: Generating Parametric Emojis with Gesture Keyboards*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, CHI EA '19 - Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, ACM Press, May 2019, p. 1-6 [*DOI :* 10.1145/3290607.3312771], https://hal.inria.fr/hal-02279111

[18] K. CARLSON, S. F. ALAOUI, G. CORNESS, T. SCHIPHORST.*Shifting Spaces: Using Defamiliarization to Design Choreographic Technologies That Support Co-Creation*, in "6th International Conference on Movement and Computing (MOCO '19)", Tempe, Arizona, United States, October 2019, n^{o} 17, p. 1–8, https://hal.inria.fr/hal-02434470

[19] A. EISELMAYER, C. WACHARAMANOTHAM, M. BEAUDOUIN-LAFON, W. MACKAY.*Touchstone2: An Interactive Environment for Exploring Trade-offs in HCI Experiment Design*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, May 2019, n^{o} 217, p. 1–11, https://hal.archives-ouvertes.fr/hal-02127273

[20] C. GRIGGIO, M. NOUWENS, J. MCGRENERE, W. MACKAY.*Augmenting Couples' Communication with Lifelines: Shared Timelines of Mixed Contextual Information*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019 [*DOI :* 10.1145/3290605.3300853], https://hal.archives-ouvertes.fr/hal-02092673

[21] S. HSUEH, S. F. ALAOUI, W. MACKAY.*Understanding Kinaesthetic Creativity in Dance*, in "ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019, n^{o} 511, 12, https://hal.archives-ouvertes.fr/hal-02430681

[22] C. KLOKMOSE, C. REMY, J. B. KRISTENSEN, R. BAGGE, M. BEAUDOUIN-LAFON, W. MACKAY.*Videostrates: Collaborative, Distributed and Programmable Video Manipulation*, in "32nd Annual ACM Symposium on User Interface Software and Technology", New Orleans, United States, Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, ACM Press, October 2019, p. 233-247 [*DOI :* 10.1145/3332165.3347912], https://hal.archives-ouvertes.fr/hal-02426650

[23] K.-D. LE, I. AVELLINO, C. FLEURY, M. FJELD, A. KUNZ.*GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration*, in "17th IFIP TC 13 International Conference on Human-Computer Interaction – INTERACT 2019", Paphos, Cyprus, Human-Computer Interaction – INTERACT 2019, September 2019, p. 282-303 [*DOI :* 10.1007/978-3-030-29384-0_18], https://hal.archives-ouvertes.fr/hal-02183386

[24] A. M. WEBB, H. FOWLER, A. KERNE, G. NEWMAN, J.-H. KIM, W. MACKAY.*Interstices: Sustained Spatial Relationships between Hands and Surfaces Reveal Anticipated Action*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019, vol. CHI 2019 [*DOI :* 10.1145/3290605.3300818], https://hal.archives-ouvertes.fr/hal-02276959

[25] A. M. WEBB, K. SPIEL, Z. O. TOUPS, B. HAMILTON, N. LUPFER, R. A. GRAEBER, W. MACKAY.*Distributed Creativity in Play*, in "ACM 2019 - 12th Creativity and Cognition Conference", San Diego, CA, United States, ACM Press, June 2019, vol. C&C 2019, p. 714-721 [*DOI :* 10.1145/3325480.3326554], https://hal.archives-ouvertes.fr/hal-02276985

[26] Y. ZHANG, N. LADEVEZE, C. FLEURY, P. BOURDOT.*Switch Techniques to Recover Spatial Consistency Between Virtual and Real World for Navigation with Teleportation*, in "16th EuroVR International Conference (EuroVR 2019)", Tallinn, Estonia, Virtual Reality and Augmented Reality - 16th EuroVR International Conference, EuroVR 2019, Springer, October 2019, vol. 11883, p. 3-23 [*DOI :* 10.1007/978-3-030-31908-3_1], https://hal.archives-ouvertes.fr/hal-02363715

National Conferences with Proceeding

[27] J. GORI, M. BEAUDOUIN-LAFON, Y. GUIARD.*Multi-evaluation of pointing and application to dyspraxia evaluation*, in "Actes de la 31e conférence francophone sur l'Interaction Homme-Machine (IHM 2019)", Grenoble, France, ACM, 2019, n^{o} 4, p. 1–12 [*DOI :* 10.1145/3366550.3372250], https://hal.archives-ouvertes.fr/hal-02381587

Conferences without Proceedings

[28] B. BRESSOLETTE, M. BEAUDOUIN-LAFON.*Flexible interfaces: future developments for post-WIMP interfaces*, in "14th International Symposium on Computer Music Multidisciplinary Research", Marseille, France, October 2019, https://hal.archives-ouvertes.fr/hal-02435177

[29] A. GOGOLOU, T. TSANDILAS, T. PALPANAS, A. BEZERIANOS.*Progressive Similarity Search on Time Series Data*, in "BigVis 2019 - 2nd International Workshop on Big Data Visual Exploration and Analytics", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-02103998

[30] W. LIU, A. OULASVIRTA, O. RIOUL, M. BEAUDOUIN-LAFON, Y. GUIARD.*Information theory: An analysis and design tool for HCI*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems - Workshop on Computational Modeling in Human-Computer Interaction", Glasgow, United Kingdom, May 2019, Position Paper, https://hal.telecom-paristech.fr/hal-02300784

Scientific Books (or Scientific Book chapters)

[31] S. F. ALAOUI.*Danse – Technologies – Phénoménologie*, in "Du signe à la performance – La notation, une pensée en mouvement", V. A. JOURNEAU, Z. CHUEKE, B. VASSILEVA (editors), L'Harmattan, June 2019, https://hal.inria.fr/hal-02434457

[32] W. MACKAY.*Designing with Sticky Notes*, in "Sticky Creativity: Post-It Note Cognition, Interaction and Digitalization", B. CHRISTENSEN, K. HALSKOV, C. KLOKMOSE (editors), Academic Press, November 2019, p. 231-256, https://hal.archives-ouvertes.fr/hal-02426827

[33] J. MALLOCH, J. GARCIA, M. M. WANDERLEY, W. MACKAY, M. BEAUDOUIN-LAFON, S. HUOT.*A Design Workbench for Interactive Music Systems*, in "New Directions in Music and Human-Computer Interaction", S. HOLLAND, T. MUDD, K. WILKIE-MCKENNA, A. MCPHERSON, M. M. WANDERLEY (editors), Springer Series on Cultural Computing, Springer, February 2019, p. 23-40 [*DOI :* 10.1007/978-3-319-92069-6_2], https://hal.inria.fr/hal-02056992

[34] M. M. WANDERLEY, W. MACKAY.*HCI, Music and Art: An Interview with Wendy Mackay*, in "New Directions in Music and Human-Computer Interaction", S. HOLLAND, T. MUDD, K. WILKIE-MCKENNA, A. MCPHERSON, M. M. WANDERLEY (editors), Springer Series on Cultural Computing, Springer, February 2019, p. 1-22 [*DOI :* 10.1007/978-3-319-92069-6_7], https://hal.inria.fr/hal-02426818

Research Reports

[35] B. CARAMIAUX, F. LOTTE, J. GEURTS, G. AMATO, M. BEHRMANN, F. BIMBOT, F. FALCHI, A. GARCIA, J. GIBERT, G. GRAVIER, H. HOLKEN, H. KOENITZ, S. LEFEBVRE, A. LIUTKUS, A. PERKIS, R. REDONDO, E. TURRIN, T. VIÉVILLE, E. VINCENT.*AI in the media and creative industries*, New European Media (NEM), April 2019, p. 1-35, https://arxiv.org/abs/1905.04175 , https://hal.inria.fr/hal-02125504

Scientific Popularization

[36] M. BEAUDOUIN-LAFON.*Le Web*, in "Sciences Numériques et Technologie, Manuel pour la classe de seconde", Editions Didier, June 2019, p. 30-53, https://hal.archives-ouvertes.fr/hal-02426663

Other Publications

[37] H. SCURTO, B. VAN KERREBROECK, B. CARAMIAUX, F. BEVILACQUA.*Designing Deep Reinforcement Learning for Human Parameter Exploration*, January 2020, https://arxiv.org/abs/1907.00824 - Preprint. Under review for publication in ACM Transactions on Computer-Human Interaction (TOCHI), https://hal.archives-ouvertes.fr/hal-02432908

References in notes

[38] A. GOGOLOU.*Iterative and Expressive Querying for Big Data Series*, Université Paris-Saclay, Orsay, France, November 2019

[39] Y. OKUYA.*CAD Modification Techniques for Design Reviews on Heterogeneous Interactive Systems*, Université Paris-Saclay, Orsay, France, November 2019

[40] P. TCHERNAVSKIJ.*Designing and Programming Malleable Software*, Université Paris-Saclay, Orsay, France, December 2019

Project-Team GAMMA

Adaptive Mesh Generation and Advanced numerical Methods

RESEARCH CENTER
Saclay - Île-de-France

THEME
Numerical schemes and simulations

Table of contents

Project-Team GAMMA

Creation of the Project-Team: 2019 June 01

Keywords:

Computer Science and Digital Science:

A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.7. - High performance computing
A6.2.8. - Computational geometry and meshes
A6.5.1. - Solid mechanics
A6.5.2. - Fluid mechanics

Other Research Topics and Application Domains:

B5.2.3. - Aviation
B5.2.4. - Aerospace
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.5.3. - Physics
B9.5.5. - Mechanics

1. Team, Visitors, External Collaborators

Research Scientists

Frederic Alauzet [Team leader, Inria, Senior Researcher, from Jun 2019, HDR]
Paul-Louis George [Inria, Senior Researcher, from Jun 2019]
Patrick Laug [Inria, Senior Researcher, from Jun 2019 until Sep 2019, HDR]
Adrien Loseille [Inria, Researcher, from Jun 2019]
David Marcum [Inria, International Chair, Advanced Research Position]

Faculty Member

Houman Borouchaki [Univ de technologie de Troyes, Professor, from Jun 2019 until Aug 2019]

External Collaborators

Loïc Frazza [Dassault Aviation, from Jun 2019]
David Marcum [MSU, from Jun 2019]
Loic Marechal [Distene, from Jun 2019]
Victorien Menier [Velo3d, from Jun 2019]

Technical Staff

Matthieu Maunoury [Inria, Engineer, from Jun 2019]
Julien Vanharen [Inria, Engineer, from Jun 2019]

PhD Students

Francesco Clerici [Inria, PhD Student, from Oct 2019]
Rémi Feuillet [École Nationale Supérieure de Techniques Avancées, PhD Student, from Jun 2019]
Lucien Rochery [Inria, PhD Student, from Sep 2019]
Lucille Marie Tenkes [Inria, PhD Student, from Jun 2019]

Administrative Assistant

Maria Agustina Ronco [Inria, Administrative Assistant, from Jun 2019]

2. Overall Objectives

2.1. Introduction

Numerical simulation has been booming over the last thirty years, thanks to increasingly powerful numerical methods, computer-aided design (CAD) and the mesh generation for complex 3D geometries, and the coming of supercomputers (HPC). The discipline is now mature and has become an integral part of design in science and engineering applications. This new status has lead scientists and engineers to consider numerical simulation of problems with ever increasing geometrical and physical complexities. A simple observation of this chart

$$\text{CAD} \longrightarrow \text{Mesh} \longrightarrow \text{Solver} \longrightarrow \text{Visualization / Analysis},$$

shows: **no mesh = no simulation** along with **"bad" mesh = wrong simulation**. We have concluded that the mesh is at the core of the classical computational pipeline and a key component to significant improvements. Therefore, the requirements on meshing methods are an ever increasing need, with increased difficulty, to produce high quality meshes to enable reliable solution output predictions in an automated manner. These requirements on meshing or equivalent technologies cannot be removed and all approaches face similar issues.

In this context, `Gamma` team was created in 1996 and has focused on the development of robust automated mesh generation methods in 3D, which was clearly a bottleneck at that time when most of the numerical simulations were 2D. The team has been very successful in tetrahedral meshing with the well-known software `Ghs3d` [53], [56] which has been distributed worldwide so far and in hexahedral meshing with the software `Hexotic` [68], [69] which was the first automated full hex mesher. The team has also worked on surface meshers with `Yams` [49] and `BLSurf` [45] and visualization with `Medit`. Before `Medit`, we were unable to visualize in real time 3D meshes !

In 2010, `Gamma3` team has replaced `Gamma` with the choice to focus more on meshing for numerical simulations. The main goal was to emphasize and to strengthen the link between meshing technologies and numerical methods (flow or structure solvers). The metric-based anisotropic mesh adaptation strategy has been very successful with the development of many error estimates, the generation of highly anisotropic meshes, its application to compressible Euler and Navier-Stokes equations [41], and its extension to unsteady problems with moving geometries [43] leading to the development of several software `Feflo.a/AMG-Lib, Wolf, Metrix, Wolf-Interpol`. A significant accomplishment was the high-fidelity prediction of the sonic boom emitted by supersonic aircraft [40]. We were the first to compute a certified aircraft sonic boom propagation in the atmosphere, thanks to mesh adaptation. The team has started to work on parallelism with the development of the multi-thread library `LPlib` and the efficient management of memory using space filling curves, and the generation of large meshes (a billion of elements) [66]. Theoretical work on high-order meshes has been also done [54].

Today, numerical simulation is an integral part of design in engineering applications with the main goal of reducing costs and speeding up the process of creating new design. Four main issues for industry are:

- Generation of a discrete surface mesh from a continuous CAD is the last non-automated step of the design pipeline and, thus, the most human time consuming
- High-performance computing (HPC) for all tools included in the design loop
- The cost in euros of a numerical simulation
- Certification of high-fidelity numerical simulations by controlling errors and uncertainties.

Let us now discuss in more details each of these issues.

Generating a discrete surface mesh from a CAD geometry definition has been the numerical analysis Achille's heel for the last 30 years. Significant issues are far too common and range from persistent translation issues between systems that can produce ill defined geometry definitions to overwhelming complexity for full configurations with all components. A geometry definition that is ill defined often does not perfectly capture the geometry's features and leads to a bad mesh and a broken simulation. Unfortunately, CAD system design is essentially decoupled from the needs of numerical simulation and is largely driven by the those of manufacturing and other areas. As a result, this step of the numerical simulation pipeline is still labor intensive and the most time consuming. There is a need to develop alternative geometry processes and models that are more suitable for numerical simulations.

Companies working on high-tech projects with high added value (Boeing, Safran, Dassault-Aviation, Ariane Group, ...) consider their design pipeline inside a HPC framework. Indeed, they are performing complex numerical simulations on complex geometries on a daily-basis, and they aim at using this in a shape-optimization loop. Therefore, any tools added to their numerical platform should be HPC compliant. This means that all developments should consider hybrid parallelism, *i.e.,* to be compatible with distributed memory architecture (MPI) and shared memory architecture (multi-threaded), to achieve scalable parallelism.

One of the main goals of numerical simulation is to reduce the cost of creating new designs (e.g reduce the number of wind-tunnel and flight tests in the aircraft industry). The emergence of 3D printers is, in some cases, making tests easier to perform, faster and cheaper. It is thus mandatory to control the cost of the numerical simulations, in other word, it is important to use less resources to achieve the same accuracy. The cost takes into account the engineer time as well as the computing resources needed to perform the numerical simulation. The cost for one simulation can vary from 15 euros for simple models (1D-2D), to 150 euros for Reynolds-averaged Navier-Stokes (3D) stationary models, or up to 15 000 euros for unsteady models like LES or Lattice-Boltzmann [0]. It is important to know that a design loop is equivalent to performing between 100 and 1 000 numerical simulations. Consequently, the need for more efficient algorithms and processes is still a key factor.

Another crucial point is checking and certification of errors and uncertainties in high-fidelity numerical simulations. These errors can come from several sources: i) modeling error (for example via turbulence models or initial conditions), ii) discretization error (due to the mesh), iii) geometry error (due to the representation of the design) and iv) implementation errors in the considered software. The error assessment and mesh generation procedure employed in the aerospace industry for CFD simulations relies heavily on the experience of the CFD user. The inadequacy of this practice even for geometries frequently encountered in engineering practice has been highlighted in studies of the AIAA [0] CFD Drag Prediction Workshops [72] and High-Lift Prediction Workshops [85], [84]. These studies suggest that the range of scales present in the turbulent flow cannot be adequately resolved using meshes generated following what is considered best present practices. In this regard, anisotropic mesh adaptation is considered as the future, as stated in the NASA report "CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences" [87] and the study dedicated to mesh adaptation [77].

These preoccupations are the core of the GAMMA project scientific program. To answer the first issue, GAMMA will focus on designing and developing a geometry modeling framework specifically intended for mesh generation and numerical simulation purposes. This is a mandatory step for automated geometry-mesh and mesh adaptation processes with an integrated geometry model. To answer the last three issues, the GAMMA team will work on the development of a high-order mesh-adaptive solution platform compatible with HPC environment. To this end, GAMMA will pursue its work on advanced mesh generation methods which should fulfill the following capabilities: i) geometric adaptive, ii) solution adaptive, iii) high-order, iv) multi-elements (structured or not), and v) using hybrid scalable parallelism. Note that items i) to iv) are based on the well-posed metric-based theoretical framework. Moreover, GAMMA will continue to work on robust flow solvers, solving the turbulent Navier-Stokes equations from second order using Finite Volume - Finite Element numerical scheme to higher-order using Flux Reconstruction (FR) method.

[0] Source Valéo and Safran Tech.

[0] The American Institute of Aeronautics and Astronautics.

The combination of adaptation - high-order - multi-elements coupled with appropriate error estimates is for the team the way to go to reduce the cost of numerical simulations while ensuring high-fidelity in a fully automated framework.

3. New Software and Platforms

3.1. Metrix

Metrix: Error Estimates and Mesh Control for Anisotropic Mesh Adaptation

KEYWORDS: Meshing - Metric - Metric fields

FUNCTIONAL DESCRIPTION: Metrix is a software that provides by various ways metric to govern the mesh generation. Generally, these metrics are constructed from error estimates (a priori or a posteriori) applied to the numerical solution. Metrix computes metric fields from scalar solutions by means of several error estimates: interpolation error, iso-lines error estimate, interface error estimate and goal oriented error estimate. It also contains several modules that handle meshes and metrics. For instance, it extracts the metric associated with a given mesh and it performs some metric operations such as: metric gradation and metric intersection.

- Participants: Adrien Loseille and Frédéric Alauzet
- Contact: Frédéric Alauzet
- URL: https://www.rocq.inria.fr/gamma/Frederic.Alauzet/code_eng.html

4. New Results

4.1. The meshing bible

Participants: Paul Louis George [correspondant], Frédéric Alauzet, Adrien Loseille, Loïc Maréchal.

Un projet important, initié en 2017, et toujours actif en 2020, consiste à écrire noir sur blanc un livre (en plusieurs volumes) et la motivation de ce travail est détaillée dans ce qui suit.

Pourquoi ce livre, pourquoi 2 volumes, pourquoi pas 3 volumes?

Notre dernier livre (généraliste) sur le maillage date de 2000 avec une mise à jour en 2008. Un collègue a commis un nouveau livre en 2015, très bien écrit mais assez classique dans son contenu, loin de préoccupations industrielles et (!) contenant quelques énormités (pas assez d'experience sur de vrais problèmes).

Ajoutons ma facilite (c'est P.L. G. qui parle) à écrire (bien ou mal, là n'est pas la question, il me suffit en effet de taper sur quelques touches d'un clavier), le désir de mon (premier) co-auteur de marquer le coup dans le domaine et la volonté (à leur corps defendant) des autres co-auteurs de participer à cette aventure. Le tout couplé avec les récents progrès dans le domaine (pensons aux éléments courbes et aux méthodes d'ordre élevé mais aussi à ce que peut être le HPC dans le domaine), tous les ingrédients sont là, on y va.

Le premier jet (un seul volume) se montre impossible à réaliser, il faudrait au minimum 800 pages, donc deux volumes a minima. Les deux volumes finis, ne reste il pas la place pour un troisième volume. Constatant avec effroi que nos étudiants (mais pas seulement) maitrisent bien force concepts mais sont incapables de voir, en pratique, comment les mettre en musique, le troisième volume est apparu comme une évidence (et on sera, au total, autour de 1000 pages).

A qui s'adresse ces volumes, bonne question. Ce n'est pas précisément de la littérature de gare mais nous nous sommes efforcé de prendre le malheureux lecteur par la main pour l'amener progressivement vers des concepts (très) avancés. Ainsi, le livre est très verbose et, en aucun cas, n'est un étalage savant de théorèmes et autres propositions, ce qui n'empêche pas de dire les choses. Par ailleurs, nous avons délibérément mis une part de subjectivité dans le propos pour suggérer (cela pouvant être contredit) que telle ou telle méthode n'avait pas notre faveur. A titre personnel, je pense que, bien que rares dans les livres, ces opinions ne peuvent qu'aider le lecteur a se former sa propre idée sur tel ou tel point.

Les livres sont publiés chez ISTE et écrits en français, eh oui, mais une traduction en anglais est avalable chez Wiley. La présence de la langue française dans la littérature scientifique me semble importante (et rejoint la politique de mon (notre) éditeur). Pour conclure, c'est plutôt satisfaisant de penser que ces livres (peut être destinés à faire référence sur le sujet) sont issus de l'Inria dans le neuf un.

4.2. Pixel-exact rendering for high-order meshes and solutions

Participants: Adrien Loseille [correspondant], Rémi Feuillet, Matthieu Maunoury.

Classic visualization software like ParaView [64], TecPlot [88], FieldView [60], Ensight [42], Medit [50], Vizir (OpenGL legacy based version) [67], Gmsh [57], ...historically rely on the display of linear triangles with linear solutions on it. More precisely, each element of the mesh is divided into a set of elementary triangles. At each vertex of the elementary triangle is attached a value and an associated color. The value and the color inside the triangle is then deduced by a linear interpolation inside the triangle. With the increase of high-order methods and high-order meshes, these softwares adapted their technology by using subdivision methods. If a mesh has high-order elements, these elements are subdivided into a set of linear triangles in order to approximate the shape of the high-order element [93]. Likewise, if a mesh has a high-order solution on it, each element is subdivided into smaller linear triangles in order to approximate the rendering of the high-order solution on it. The subdivision process can be really expensive if it is done in a naive way. For this reason, mesh adaptation procedures [80], [70], [71] are used to efficiently render high-order solutions and high-order elements using the standard linear rendering approaches. Even when optimized these approaches do have a huge RAM memory footprint as the subdivision is done on CPU in a preprocessing step. Also the adaptive subdivision process can be dependent on the palette (*i.e.* the range of values where the solution is studied) as the color only vary when the associated value is in this range. In this case, a change of palette inevitably imposes a new adaptation process. Finally, the use of a non conforming mesh adaptation can lead to a discontinuous rendering for a continuous solution.

Other approaches are specifically devoted to high-order solutions and are based on ray casting [75], [76], [78]. The idea is for a given pixel, to find exactly its color. To do so, for each pixel, rays are cast from the position of the screen in the physical space and their intersection with the scene determines the color for the pixel. If high-order features are taken into account, it determines the color exactly for this pixel. However, this method is based on two non-linear problems: the root-finding problem and the inversion of the geometrical mapping. These problems are really costly and do not compete with the interactivity of the standard linear rendering methods even when these are called with a subdivision process unless they are done conjointly on the GPU. However, synchronization between GPU and `OpenGL` buffer are non-trivial combination.

The proposed method intends to be a good compromise between both methods. It does guarantee pixel-exact rendering on linear elements without extra subdivision or ray casting and it keeps the interactivity of a classical method. Moreover, the subdivision of the curved entities is done on the fly on GPU which leaves the RAM memory footprint at the size of the loaded mesh.

4.3. High-order mesh generation

Participants: Frédéric Alauzet [correspondant], Adrien Loseille, Rémi Feuillet, Dave Marcum, Lucien Rochery.

For years, the resolution of numerical methods has consisted in solving Partial Derivative Equations by means of a piecewise linear representation of the physical phenomenon on linear meshes. This choice was merely driven by computational limitations. With the increase of the computational capabilities, it became possible to increase the polynomial order of the solution while keeping the mesh linear. This was motivated by the fact that even if the increase of the polynomial order requires more computational resources per iteration of the solver, it yields a faster convergence of the approximation error [0] [92] and it enables to keep track of unsteady features for a longer time and with a coarser mesh than with a linear approximation of the solution. However, in [46], [65], it was theoretically shown that for elliptic problems the optimal convergence rate for a high-order

[0] The order of convergence is the degree of the polynomial approximation plus one.

method was obtained with a curved boundary of the same order and in [44], evidence was given that without a high-order representation of the boundary the studied physical phenomenon was not exactly solved using a high-order method. In [95], it was even highlighted that, in some cases, the order of the mesh should be of a higher degree than the one of the solver. In other words, if the used mesh is not a high-order mesh, then the obtained high-order solution will never reliably represent the physical phenomenon.

Based on these issues, the development of high-order mesh generation procedures appears mandatory. To generate high-order meshes, several approaches exist. The first approach was tackled twenty years ago [47] for both surface and volume meshing. At this moment the idea was to use all the meshing tools to get a valid high-order mesh. The same problem was revisited a few years later in [86] for bio-medical applications. In these first approaches and in all the following, the underlying idea is to use a linear mesh and elevate it to the desired order. Some make use of a PDE or variational approach to do so [39], [79], [48], [73], [91], [94], [58], others are based on optimization and smoothing operations and start from a linear mesh with a constrained high-order curved boundary in order to generate a suitable high-order mesh [62], [51], [89]. Also, when dealing with Navier-Stokes equations, the question of generating curved boundary layer meshes (also called viscous meshes) appears. Most of the time, dedicated approaches are set-up to deal with this problem [74], [63]. In all these techniques, the key feature is to find the best deformation to be applied to the linear mesh and to optimize it. The prerequisite of these methods is that the initial boundary is curved and will be used as an input data. A natural question is consequently to study an optimal position of the high-order nodes on the curved boundary starting from an initial linear or high-order boundary mesh. This can be done in a coupled way with the volume [81], [90] or in a preprocessing phase [82], [83]. In this process, the position of the nodes is set by projection onto the CAD geometry or by minimization of an error between the surface mesh and the CAD surface. Note that the vertices of the boundary mesh can move as well during the process. In the case of an initial linear boundary mesh with absence of a CAD geometry, some approaches based on normal reconstructions can be used to create a surrogate for the CAD model [93], [59]. Finally, a last question remains when dealing with such high-order meshes: Given a set of degrees of freedom, is the definition of these objects always valid ?. Until the work presented in [55], [61], [52], no real approach was proposed to deal in a robust way with the validity of high-order elements. The novelty of these approaches was to see the geometrical elements and their Jacobian as Bézier entities. Based on the properties of the Bézier representation, the validity of the element is concluded in a robust sense, while the other methods were only using a sampling of the Jacobian to conclude about its sign without any warranty on the whole validity of the elements.
In this context, several issues have been addressed : the analogy between high-order and Bézier elements, the development of high-order error estimates suitable for parametric high-order surface mesh generation andthe generalization of mesh optimization operators and their applications to curved mesh generation, moving-mesh methods, boundary layer mesh generation and mesh adaptation.

4.4. Unstructured anisotropic mesh adaptation for 3D RANS turbomachinery applications

Participants: Frédéric Alauzet, Loïc Frazza, Adrien Loseille [correspondant], Julien Vanharen.

The scope of this paper is to demonstrate the viability and efficiency of unstructured anisotropic mesh adaptation techniques to turbomachinery applications. The main difficulty in turbomachinery is the periodicity of the domain that must be taken into account inthe solution mesh-adaptive process. The periodicity is strongly enforced in the flow solver using ghost cells to minimize the impact on the source code. For the mesh adaptation, the local remeshing is done in two steps. First, the inner domain is remeshed with frozen periodic frontiers, and, second, the periodic surfaces are remeshed after moving geometrice ntities from one side of the domain to the other. One of the main goal of this work is to demonstrate how mesh adaptation, thanks to its automation, is able to generate meshes that are extremely difficult to envision and almost impossible to generate manually. This study only considers feature-based error estimate based on the standard multi-scale Lpinterpolation error estimate. We presents all the specific modifications that have been introduced in the adaptive process to deal with periodic simulations used for turbomachinery applications. The periodic mesh adaptation strategy is then tested and validated on the LS89 high pressure axial turbine vane and the NASA Rotor 37 test cases.

4.5. Hybrid mesh adaptation for CFD simulations

Participants: Frédéric Alauzet [correspondant], Lucille Tenkès, Julien Vanharen.

The aim of mesh adaptation is to generate the optimal mesh to perform a specific numerical simulation. It is nowadays a mature tool which is mathematically well-posed and fully automatic regarding tetrahedral meshes. Yet, there is still a strong demand for structured meshes, as many numerical schemes have proven to be more accurate on quadrilateral meshes than on triangular meshes, and as many favor structured elements in the boundary layer instead of tetrahedra to simulate viscous turbulent flows. Since no method can automatically provide pure hexahedral adapted meshes respecting alignment constraints, one solution is to use hybrid meshes, *i.e.* meshes containing both structured and unstructured elements. Accordingly, the following work focuses on hybrid metric-based mesh adaptation and CFD simulation on such meshes. Regarding hybrid mesh generation, the method relies on a preliminary mesh obtained through so-called metric-aligned and metric-orthogonal approaches. These approaches utilize the directional information held by a prescribed metric-field to generate right angled elements, that can be combined into structured elements to form a hybrid mesh. The result highly depends on the quality of the metric field. Thus, emphasis is put on the size gradation control performed beforehand. This process is re-designed to favor metric-orthogonal meshes. To validate the method, some CFD simulations are performed. The modifications brought to the existing Finite Volume solver to enable such computations has been developed.

4.6. Anisotropic mesh adaptation for fluid-structure interactions

Participants: Frédéric Alauzet, Adrien Loseille, Julien Vanharen [correspondant].

A new strategy for mesh adaptation dealing with Fluid-Structure Interaction (FSI) problems is presented using a partitioned approach. The Euler equations are solved by an edge-based Finite Volume solver whereas the linear elasticity equations are solved by the Finite Element Method using the Lagrange P^1 elements. The coupling between both codes is realized by imposing boundary conditions. Small displacements of the structure are assumed and so the mesh is not deformed. The computation of a well-documented FSI test case is finally carried out to perform validation of this new strategy.

5. Bilateral Contracts and Grants with Industry

5.1. Bilateral Contracts with Industry

- Boeing
- Safran Tech

5.2. Bilateral Grants with Industry

- Projet RAPID DGA

6. Partnerships and Cooperations

6.1. National Initiatives

6.1.1. ANR

6.1.1.1. ANR IMPACTS 2018-2021

Ideal Mesh generation for modern solvers and comPuting ArchiteCTureS.

- Coordinateur : Adrien Loseille

- The rapid improvement of computer hardware and physical simulation capabilities has revolutionized science and engineering, placing computational simulation on an equal footing with theoretical analysis and physical experimentation. This rapidly increasing reliance on the predictive capabilities has created the need for rigorous control of numerical errors which strongly impact these predictions. A rigorous control of the numerical error can be only achieved through mesh adaptivity. In this context, the role of mesh adaptation is prominent, as the quality of the mesh, its refinement, and its alignment with the physics are major contributions to these numerical errors. The IMPACTS project aims at pushing the envelope in mesh adaptation in the context of large size, very high fidelity simulations by proposing a new adaptive mesh generation framework. This framework will be based on new theoretical developments on Riemannian metric-field and on innovative algorithmic developments coupling a unique cavity-operator with an advancing-point techniques in order to produce high quality hybrid, curved and adapted meshes.

7. Bibliography

Major publications by the team in recent years

[1] F. ALAUZET, A. LOSEILLE.*High Order Sonic Boom Modeling by Adaptive Methods*, in "Journal Of Computational Physics", 2010, vol. 229, p. 561-593, http://dx.doi.org/10.1016/j.jcp.2009.09.020

[2] F. ALAUZET, M. MEHRENBERGER.*P1-conservative solution interpolation on unstructured triangular meshes*, Inria, January 2009, http://hal.archives-ouvertes.fr/inria-00354509/

[3] T. AMARI, A. CANOU, J.-J. ALY, F. DELYON, F. ALAUZET.*Magnetic cage and rope as the key for solar eruptions*, in "Nature", February 2018, vol. 554, n° 7691, p. 211 - 215 [*DOI :* 10.1038/NATURE24671], https://hal-cea.archives-ouvertes.fr/cea-01872913

[4] A. BELME, F. ALAUZET, A. DERVIEUX.*An a priori anisotropic Goal-Oriented Error Estimate for Viscous Compressible Flow and Application to Mesh Adaptation*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01927113

[5] H. BOROUCHAKI, D. CHAPELLE, P.-L. GEORGE, P. LAUG, P.-J. FREY.*Estimateurs d'erreur géométriques et adaptation de maillages*, in "Maillage et adaptation ; Traité MIM, série Méthodes numériques et éléments finis", Hermès, Paris, France, 2001, p. 279-310

[6] Y. BOURGAULT, M. PICASSO, F. ALAUZET, A. LOSEILLE.*On the use of anisotropic error estimators for the adaptative solution of 3-D inviscid compressible flows*, in "International Journal for Numerical Methods in Fluids", 2009, vol. 59, p. 47-74, http://onlinelibrary.wiley.com/doi/10.1002/fld.1797/abstract

[7] S. CHAILLAT, S. P. GROTH, A. LOSEILLE.*Metric-based anisotropic mesh adaptation for 3D acoustic boundary element methods*, in "Journal of Computational Physics", November 2018, vol. 372, p. 473 - 499 [*DOI :* 10.1016/J.JCP.2018.06.048], https://hal.archives-ouvertes.fr/hal-01895636

[8] A. DERVIEUX, E. GAUCI, L. FRAZZA, A. BELME, A. CARABIAS, A. LOSEILLE, F. ALAUZET.*Mesh-Anpassung für k-genaue Approximationen in CFD*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01927145

[9] R. FEUILLET, A. LOSEILLE, F. ALAUZET.*P2 mesh optimization operators*, in "27th International Meshing Roundtable", Albuquerque, United States, October 2018, https://hal.inria.fr/hal-01962132

[10] R. FEUILLET, A. LOSEILLE, D. MARCUM, F. ALAUZET.*Connectivity-change moving mesh methods for high-order meshes: Toward closed advancing-layer high-order boundary layer mesh generation*, in "2018 Fluid Dynamics Conference, AIAA AVIATION Forum", Atlanta, United States, June 2018 [*DOI :* 10.2514/6.2018-4167], https://hal.inria.fr/hal-01962129

[11] L. FRAZZA.*3D anisotropic mesh adaptation for Reynolds Averaged Navier-Stokes simulations*, Sorbonne Université , UPMC, December 2018, https://hal.inria.fr/tel-01962318

[12] L. FRAZZA, A. LOSEILLE, F. ALAUZET, A. DERVIEUX.*Nonlinear corrector for RANS equations*, in "2018 Fluid Dynamics Conference", June 2018 [*DOI :* 10.2514/6.2018-3242], https://hal.inria.fr/hal-01927129

[13] E. GAUCI, A. BELME, A. CARABIAS, A. LOSEILLE, F. ALAUZET, A. DERVIEUX.*A priori error-based mesh adaptation in CFD*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01928249

[14] P.-L. GEORGE, H. BOROUCHAKI, P.-J. FREY, P. LAUG, E. SALTEL.*17*, in "Mesh Generation and Mesh Adaptivity: Theories and Techniques", Wiley InterScience, 2004, p. 497-523, ISBN 0-470-84699-2, E. Stein, R. de Borst and T.J.R. Hughes ed., 2nd edition 2008

[15] H. GUILLARD, J. LAKHLILI, A. LOSEILLE, A. LOYER, B. NKONGA, A. RATNANI, A. ELARIF.*Tokamesh : A software for mesh generation in Tokamaks*, CASTOR, December 2018, n° RR-9230, https://hal.inria.fr/hal-01948060

[16] J. V. LANGENHOVE, D. LUCOR, F. ALAUZET, A. BELME.*Goal-oriented error control of stochastic system approximations using metric-based anisotropic adaptations*, February 2018, https://arxiv.org/abs/1805.00370 - working paper or preprint, https://hal.inria.fr/hal-01703054

[17] A. LOSEILLE, F. ALAUZET.*Optimal 3D Highly Anisotropic Mesh Adaptation based on the Continuous Mesh Framework*, in "18th International meshing roundtable", Springer, 2009, http://dx.doi.org/10.1007/978-3-642-04319-2_20

[18] A. LOSEILLE, A. DERVIEUX, F. ALAUZET.*Fully anisotropic goal-oriented mesh adaptation for 3D steady Euler equations*, in "Journal Of Computational Physics", 2010, vol. 229, p. 2866-2897, http://dx.doi.org/10.1016/j.jcp.2009.12.021

[19] A. LOSEILLE, R. FEUILLET.*Vizir: High-order mesh and solution visualization using OpenGL 4.0 graphic pipeline*, in "2018 - AIAA Aerospace Sciences Meeting, AIAA SciTech Forum", kissimmee, United States, January 2018, p. 1-13 [*DOI :* 10.2514/6.2018-1174], https://hal.inria.fr/hal-01686714

[20] A. LOSEILLE, R. LÖHNER.*Robust Boundary Layer Mesh Generation*, in "Proceedings of the 21st International Meshing Roundtable", X. JIAO, J.-C. WEILL (editors), Springer Berlin Heidelberg, 2012, p. 493-511 [*DOI :* 10.1007/978-3-642-33573-0_29], http://hal.inria.fr/hal-00935315

[21] E. MONTREUIL, W. GHEDHAIFI, V. CHMIELARSKI, V. FRANÇOIS, F. GAND, A. LOSEILLE.*Numerical Simulation of contrail formation on the Common Research Model wing/body/engine configuration*, in "AIAA Aviation and Aeronautics Forum and Exposition 2018", Atlanta, United States, June 2018, https://hal.archives-ouvertes.fr/hal-01961143

[22] J. TOMASI, B. MENNUCCI, P. LAUG.*The modeling and simulation of the liquid phase*, in "Handbook of Numerical Analysis, Volume X, Special Volume on Computational Chemistry", P.-G. CIARLET, C. LE BRIS (editors), North-Holland, Amsterdam, Netherlands, 2003, p. 271-375, ISBN: 0-444-51248-9

Publications of the year

Articles in International Peer-Reviewed Journal

[23] F. AMLANI, S. CHAILLAT, A. LOSEILLE.*An efficient preconditioner for adaptive Fast Multipole accelerated Boundary Element Methods to model time-harmonic 3D wave propagation*, in "Computer Methods in Applied Mechanics and Engineering", 2019, vol. 352, n^o 1, p. 189-210 [*DOI :* 10.1016/J.CMA.2019.04.026], https://hal.archives-ouvertes.fr/hal-02113702

[24] S. ANTONY, A. CHEROUAT, G. MONTAY.*Hemp fibre woven fabrics / polypropylene based honeycomb sandwich structure for aerospace applications*, in "Advances in Aircraft and Spacecraft Science", March 2019, vol. 6, n^o 2, p. 87-103 [*DOI :* 10.12989/AAS.2019.6.2.087], https://hal-utt.archives-ouvertes.fr/hal-02278252

[25] D. BARCHIESI, D. CAKIR, T. GROSGES, N. FRÉTY, E. ANGLARET.*Recovering effective thicknesses and optical properties of copper and copper oxide layers from absorbance measurements*, in "Optical Materials", May 2019, vol. 91, p. 138-146 [*DOI :* 10.1016/J.OPTMAT.2019.02.029], https://hal.inria.fr/hal-02077581

[26] B. SAIDI, L. GIRAUD MOREAU, S. MHEMED, A. CHEROUAT, P.-A. ADRAGNA, R. NASRI.*Hot incremental forming of titanium human skull prosthesis by using cartridge heaters: a reverse engineering approach*, in "International Journal of Advanced Manufacturing Technology", March 2019, vol. 101, n^o 1-4, p. 873-880 [*DOI :* 10.1007/S00170-018-2975-9], https://hal-utt.archives-ouvertes.fr/hal-02273283

[27] S. ZHU, C. BOUBY, A. CHEROUAT, T. BEN ZINEB.*3D reconstitution and numerical analysis of superelastic behavior of porous shape memory alloy*, in "International Journal of Solids and Structures", August 2019, vol. 168, p. 109-122 [*DOI :* 10.1016/J.IJSOLSTR.2019.03.023], https://hal-utt.archives-ouvertes.fr/hal-02272260

International Conferences with Proceedings

[28] S. ANTONY, A. CHEROUAT, G. MONTAY.*Multiscale analysis to investigate the mechanical and forming behaviour of hemp fibre woven fabrics / polypropylene composite*, in "22nd International Conference on Composite Materials", Melbourne, Australia, ICCM22 2019. Melbourne, VIC: Engineers Australia, 2019, p. 318-327, https://hal-utt.archives-ouvertes.fr/hal-02473963

[29] R. FEUILLET, O. COULAUD, A. LOSEILLE.*Anisotropic Error Estimate for High-order Parametric Surface Mesh Generation*, in "28th International Meshing Roundtable", Buffalo, NY, United States, October 2019, https://hal.inria.fr/hal-02345068

[30] R. FEUILLET, D. MARCUM, F. ALAUZET.*A closed advancing-layer method for generating curved boundary layer mesh*, in "AIAA Aviation 2019 Forum", Dallas, TX, United States, American Institute of Aeronautics and Astronautics, June 2019 [*DOI :* 10.2514/6.2019-3675], https://hal.inria.fr/hal-02345334

[31] W. GHEDHAIFI, A. BIENNER, R. MEGHERBI, E. MONTREUIL, E. TERRENOIRE, X. VANCASSEL, A. LOSEILLE.*Influence of atmospheric conditions on contrail formation: 3D simulation versus Schmidt-*

Appleman criterion, in "ISABE 2019 - 24th ISABE Conference", Canberra, Australia, September 2019, https://hal.archives-ouvertes.fr/hal-02470854

[32] R. HU, S. BOUINDOUR, H. SNOUSSI, A. CHEROUAT, C. CHAHLA.*Multiple Cues Association for Multiple Object Tracking Based on Convolutional Neural Network*, in "IEEE AIKE 2019 - Second International Conference on Artificial Intelligence and Knowledge Engineering", Cagliari, Italy, 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), IEEE, June 2019, p. 117-122 [*DOI :* 10.1109/AIKE.2019.00030], https://hal-utt.archives-ouvertes.fr/hal-02353268

Conferences without Proceedings

[33] F. GUIBAULT, H. BOROUCHAKI, P. LAUG.*Efficient high-order discrete geometric model from CAD*, in "NAHOMCon19 - North American High Order Method conference", San Diego, CA, United States, June 2019, https://hal.inria.fr/hal-02360842

[34] A. HALOUANI, M. M. CHAABANE, M. HADDAR, A. CHEROUAT.*Effect of Harmonic Excitation on PCB and Component Assembly*, in "Advances in Acoustics and Vibration II", Hammamet, Tunisia, September 2019, p. 149-154, https://hal-utt.archives-ouvertes.fr/hal-02280428

[35] A. HALOUANI, A. CHEROUAT, M. M. CHAABANE, M. HADDAR.*A Probabilistic approach to the robust thermo-mechanical analysis of Ball Grid Array Solder Joints*, in "EuroSimE 2019 - 20th International Conference in Thermal, Mechanical & Multiphysics Simulation and Experiments in Micro/Nano-Electronics and Micro/Nano-systems", Hannover, Germany, IEEE, March 2019, p. 1-6, https://hal-utt.archives-ouvertes.fr/hal-02278339

Scientific Books (or Scientific Book chapters)

[36] A. LIMARE, H. BOROUCHAKI, P. BRENNER.*Adaptive Mesh Refinement with an Automatic Hybrid RANS/LES Strategy and Overset Grids*, in "Progress in Hybrid RANS-LES Modelling", November 2019, p. 159-168 [*DOI :* 10.1007/978-3-030-27607-2_12], https://hal-utt.archives-ouvertes.fr/hal-02378168

Research Reports

[37] R. FEUILLET, A. LOSEILLE, F. ALAUZET.*Mesh adaptation for the embedded boundary method in CFD*, Inria Saclay - Ile-de-France, November 2019, n° RR-9305 [*DOI :* 10.13140/RG.2.2.22343.44969], https://hal.inria.fr/hal-02378738

Other Publications

[38] R. COGRANNE, R. SLYSZ, L. MOREAU, H. BOROUCHAKI.*A new Edge Detector Based on Parametric Surface Model: Regression Surface Descriptor*, September 2019, https://arxiv.org/abs/1904.10235 - 21 pages, 13 figures and 2 tables, https://hal-utt.archives-ouvertes.fr/hal-02290249

References in notes

[39] R. ABGRALL, C. DOBRZYNSKI, A. FROEHLY.*A method for computing curved meshes via the linear elasticity analogy, application to fluid dynamics problems*, in "International Journal for Numerical Methods in Fluids", 2014, vol. 76, n° 4, p. 246–266

[40] F. ALAUZET, A. LOSEILLE.*High Order Sonic Boom Modeling by Adaptive Methods*, in "J. Comp. Phys.", 2010, vol. 229, p. 561-593

[41] F. ALAUZET, A. LOSEILLE.*A decade of progress on anisotropic mesh adaptation for Computational Fluid Dynamics*, in "Comput. Aided Des.", 2016, vol. 72, p. 13-39

[42] ANSYS INC..*Ensight*, https://www.ansys.com/en/products/platform/ansys-ensight

[43] N. BARRAL, G. OLIVIER, F. ALAUZET.*Metric-based anisotropic mesh adaptation for three-dimensional time-dependent problems involving moving geometries*, in "J. Comp. Phys.", 2017, vol. 331, p. 157-187

[44] F. BASSI, S. REBAY.*High-order accurate discontinuous finite element solution of the 2D Euler equations*, in "Journal of computational physics", 1997, vol. 138, n^o 2, p. 251–285

[45] H. BOROUCHAKI, P. LAUG, P. GEORGE.*Parametric surface meshing using a combined advancing-front – generalized-Delaunay approach*, in "Int. J. Numer. Meth. Engng", 2000, vol. 49, n^o 1-2, p. 233-259

[46] P. G. CIARLET, P.-A. RAVIART.*The combined effect of curved boundaries and numerical integration in isoparametric finite element methods*, in "The mathematical foundations of the finite element method with applications to partial differential equations", Elsevier, 1972, p. 409–474

[47] S. DEY, R. M. O'BARA, M. S. SHEPHARD.*Curvilinear Mesh Generation in 3D*, in "Proceedings of the 7th International Meshing Roundtable", 1999, p. 407–417

[48] M. FORTUNATO, P.-O. PERSSON.*High-order Unstructured Curved Mesh Generation Using the Winslow Equations*, in "J. Comput. Phys.", February 2016, vol. 307, p. 1–14

[49] P. FREY.*About surface remeshing*, in "Proceedings of the 9th International Meshing Roundtable", New Orleans, LO, USA, 2000, p. 123-136

[50] P. J. FREY.*Medit: An interactive mesh visualization software, Inria Technical Report RT0253*, 2001

[51] A. GARGALLO-PEIRÓ, X. ROCA, J. PERAIRE, J. SARRATE.*Defining quality measures for mesh optimization on parameterized CAD surfaces*, in "Proceedings of the 21st International Meshing Roundtable", Springer, 2013, p. 85–102

[52] P. L. GEORGE, H. BOROUCHAKI, N. BARRAL.*Geometric validity (positive jacobian) of high-order Lagrange finite elements, theory and practical guidance*, in "Engineering with computers", 2016, vol. 32, n^o 3, p. 405–424

[53] P. GEORGE, H. BOROUCHAKI.*"Ultimate" robustness in meshing an arbitrary polyhedron*, in "Int. J. Numer. Meth. Engng", 2003, vol. 58, n^o 7, p. 1061-1089

[54] P. GEORGE, H. BOROUCHAKI.*Construction of tetrahedral meshes of degree two*, in "Int. J. Numer. Meth. Engng", 2012, vol. 90, p. 1156-1182

[55] P. L. GEORGE, H. BOROUCHAKI.*Construction of tetrahedral meshes of degree two*, in "International Journal for Numerical Methods in Engineering", 2012, vol. 90, n^o 9, 1156,1182

[56] P. GEORGE, F. HECHT, E. SALTEL.*Automatic mesh generator with specified boundary*, in "Comput. Methods Appl. Mech. Engrg.", 1991, vol. 92, p. 269-288

[57] C. GEUZAINE, J.-F. REMACLE.*Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities*, in "International Journal for Numerical Methods in Engineering", 2009, vol. 79, n^{o} 11, p. 1309-1331

[58] R. HARTMANN, T. LEICHT.*Generation of unstructured curvilinear grids and high-order discontinuous Galerkin discretization applied to a 3D high-lift configuration*, in "International Journal for Numerical Methods in Fluids", 2016, vol. 82, n^{o} 6, p. 316-333

[59] J. IMS, Z. J. WANG.*Automated low-order to high-order mesh conversion*, in "Engineering with Computers", Jan 2019, vol. 35, n^{o} 1, p. 323–335

[60] INTELLIGENT LIGHT.*FieldView*, http://www.ilight.com/en/products/fieldview-18

[61] A. JOHNEN, J.-F. REMACLE, C. GEUZAINE.*Geometrical validity of curvilinear finite elements*, in "Journal of Computational Physics", 2013, vol. 233, n^{o} 15, p. 359-372

[62] S. L. KARMAN, J. T. ERWIN, R. S. GLASBY, D. STEFANSKI.*High-Order Mesh Curving Using WCN Mesh Optimization*, in "46th AIAA Fluid Dynamics Conference", AIAA AVIATION Forum, American Institute of Aeronautics and Astronautics, 2016

[63] S. L. KARMAN.*Curving for Viscous Meshes*, in "27th International Meshing Roundtable", Cham, Springer International Publishing, 2019, p. 303–325

[64] KITWARE INC..*ParaView*, https://www.paraview.org/

[65] M. LENOIR.*Optimal isoparametric finite elements and error estimates for domains involving curved boundaries*, in "SIAM journal on numerical analysis", 1986, vol. 23, n^{o} 3, p. 562–580

[66] A. LOSEILLE, F. ALAUZET, V. MENIER.*Unique cavity-based operator and hierarchical domain partitioning for fast parallel generation of anisotropic meshes*, in "Comput. Aided Des.", 2017, vol. 85, p. 53-67

[67] A. LOSEILLE, H. GUILLARD, A. LOYER.*An introduction to Vizir: an interactive mesh visualization and modification software*, 2016, EOCOE, Rome, Italy

[68] L. MARÉCHAL.*A new approach to octree-based hexahedral meshing*, in "Proceedings of the 10th International Meshing Roundtable", Newport Beach, CA, USA, 2001

[69] L. MARÉCHAL.*Advances in Octree-Based All-Hexahedral Mesh Generation: Handling Sharp Features*, in "Proceedings of the 18th International Meshing Roundtable", Salt Lake City, UT, USA, 2009, p. 65-84

[70] M. MAUNOURY, C. BESSE, V. MOUYSSET, S. PERNET, P.-A. HAAS.*Well-suited and adaptive post-processing for the visualization of hp simulation results*, in "Journal of Computational Physics", 2018, vol. 375, p. 1179 - 1204

[71] M. MAUNOURY.*Méthode de visualisation adaptée aux simulations d'ordre élevé. Application à la compression-reconstruction de champs rayonnés pour des ondes harmoniques.*, Université de Toulouse, February 2019

[72] D. MAVRIPLIS.*Results from the 3rd Drag Prediction Workshop using NSU3D unstructured mesh solver*, in "45th AIAA Aerospace Sciences Meeting", AIAA-2007-0256, Reno, NV, USA, Jan 2007

[73] D. MOXEY, D. EKELSCHOT, Ü. KESKIN, S. SHERWIN, J. PEIRÒ.*High-order curvilinear meshing using a thermo-elastic analogy*, in "Computer-Aided Design", 2016, vol. 72, p. 130 - 139

[74] D. MOXEY, M. GREEN, S. SHERWIN, J. PEIRÓ.*An isoparametric approach to high-order curvilinear boundary-layer meshing*, in "Computer Methods in Applied Mechanics and Engineering", 2015, vol. 283, p. 636 - 650

[75] B. NELSON, R. HAIMES, R. M. KIRBY.*GPU-Based Interactive Cut-Surface Extraction From High-Order Finite Element Fields*, in "IEEE Transactions on Visualization and Computer Graphics", 2011, vol. 17, n° 12, p. 1803–11

[76] B. NELSON, E. LIU, R. M. KIRBY, R. HAIMES.*ElVis: A System for the Accurate and Interactive Visualization of High-Order Finite Element Solutions*, in "IEEE Transactions on Visualization and Computer Graphics", 2012, vol. 18, n° 12, p. 2325-2334

[77] M. PARK, J. KRAKOS, T. MICHAL, A. LOSEILLE, J. ALONSO.*Unstructured Grid Adaptation: Status, Potential Impacts, and Recommended Investments Toward CFD Vision 2030*, in "46th AIAA Fluid Dynamics Conference", AIAA Paper 2016-3323, Washington, D.C., USA, 2016

[78] J. PEIRÓ, D. MOXEY, B. JORDI, S. J. SHERWIN, B. NELSON, R. M. KIRBY, R. HAIMES.*High-Order Visualization with ElVis*, in "Notes on Numerical Fluid Mechanics and Multidisciplinary Design", Springer International Publishing, 2015, p. 521–534

[79] P.-O. PERSSON, J. PERAIRE.*Curved mesh generation and mesh refinement using Lagrangian solid mechanics*, in "47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition", 2009, 949

[80] J.-F. REMACLE, N. CHEVAUGEON, E. MARCHANDISE, C. GEUZAINE.*Efficient visualization of high-order finite elements*, in "International Journal for Numerical Methods in Engineering", 2007, vol. 69, n° 5, p. 750-771

[81] E. RUIZ-GIRONÈS, X. ROCA, J. SARRATE.*High-order mesh curving by distortion minimization with boundary nodes free to slide on a 3D CAD representation*, in "Computer-Aided Design", 2016, vol. 72, p. 52 - 64, 23rd International Meshing Roundtable Special Issue: Advances in Mesh Generation

[82] E. RUIZ-GIRONÈS, J. SARRATE, X. ROCA.*Defining an L^2-disparity Measure to Check and Improve the Geometric Accuracy of Non-interpolating Curved High-order Meshes*, in "Procedia Engineering", 2015, vol. 124, p. 122–134

[83] E. RUIZ-GIRONÈS, J. SARRATE, X. ROCA.*Generation of curved high-order meshes with optimal quality and geometric accuracy*, in "Procedia engineering", 2016, vol. 163, p. 315–327

[84] C. RUMSEY, J. SLOTNICK, M. LONG, R. STUEVER, T. WAYMAN.*Summary of the first AIAA CFD High-Lift Prediction Workshop*, in "Journal of Aircraft", 2011, vol. 48, n° 6, p. 2068-2079

[85] C. RUMSEY, J. SLOTNICK.*Overview and Summary of the Second AIAA High Lift Prediction Workshop*, in "Journal of Aircraft", 2015, vol. 52, n^{o} 4, p. 1006-1025

[86] S. J. SHERWIN, J. PEIRÓ.*Mesh generation in curvilinear domains using high-order elements*, in "International Journal for Numerical Methods in Engineering", 2002, vol. 53, n^{o} 1, p. 207-223

[87] J. SLOTNICK, A. KHODADOUST, J. ALONSO, D. DARMOFAL, W. GROPP, E. LURIE, D. MAVRIPLIS.*CFD Vision 2030 Study: A path to revolutionary computational aerosciences*, NASA, March 2014

[88] TECPLOT INC..*TecPlot*, https://www.tecplot.com/

[89] T. TOULORGE, C. GEUZAINE, J.-F. REMACLE, J. LAMBRECHTS.*Robust untangling of curvilinear meshes*, in "Journal of Computational Physics", 2013, vol. 254, p. 8 - 26

[90] T. TOULORGE, J. LAMBRECHTS, J.-F. REMACLE.*Optimizing the geometrical accuracy of curvilinear meshes*, in "Journal of Computational Physics", 2016, vol. 310, p. 361 - 380

[91] M. TURNER, J. PEIRÒ, D. MOXEY.*A Variational Framework for High-order Mesh Generation*, in "Procedia Engineering", 2016, vol. 163, n^{o} Supplement C, p. 340 - 352, 25th International Meshing Roundtable

[92] J. VANHAREN.*High-order numerical methods for unsteady flows around complex geometries*, Université de Toulouse, 2017

[93] A. VLACHOS, P. JÖRG, C. BOYD, J. L. MITCHELL.*Curved PN Triangles*, in "Proceedings of the 2001 Symposium on Interactive 3D Graphics", 2001, p. 159-166

[94] Z. Q. XIE, R. SEVILLA, O. HASSAN, K. MORGAN.*The generation of arbitrary order curved meshes for 3D finite element analysis*, in "Computational Mechanics", Mar 2013, vol. 51, n^{o} 3, p. 361–374

[95] P. ZWANENBURG, S. NADARAJAH.*On the Necessity of Superparametric Geometry Representation for Discontinuous Galerkin Methods on Domains with Curved Boundaries*, in "23rd AIAA Computational Fluid Dynamics Conference", 2017

Project-Team GRACE

Geometry, arithmetic, algorithms, codes and encryption

IN COLLABORATION WITH: Laboratoire d'informatique de l'école polytechnique (LIX)

IN PARTNERSHIP WITH:
CNRS

Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Algorithmics, Computer Algebra and Cryptology

Table of contents

Project-Team GRACE

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

A1.2.5. - Internet of things
A1.2.8. - Network security
A1.3.3. - Blockchain
A2.3.1. - Embedded systems
A4.2. - Correcting codes
A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.3. - Cryptographic protocols
A4.3.4. - Quantum Cryptography
A4.4. - Security of equipment and software
A4.8. - Privacy-enhancing technologies
A4.9. - Security supervision
A8.1. - Discrete mathematics, combinatorics
A8.4. - Computer Algebra
A8.5. - Number theory

Other Research Topics and Application Domains:

B5.11. - Quantum systems
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Daniel Augot [Team leader, Inria, Senior Researcher, HDR]
Alain Couvreur [Inria, Researcher, HDR]
Benjamin Smith [Inria, Researcher]

Faculty Members

Françoise Levy-Dit-Vehel [École Nationale Supérieure de Techniques Avancées, Professor, HDR]
François Morain [École polytechnique, Professor, HDR]
Guénaël Renault [École polytechnique, Professor, HDR]

External Collaborators

Elise Barelli [Univ de Versailles Saint-Quentin-en-Yvelines]
Luca de Feo [Univ de Versailles Saint-Quentin-en-Yvelines]
Julien Lavauzelle [Univ de Rennes I]
Philippe Lebacque [Univ de Franche-Comté, until Aug 2019]
Matthieu Rambaud [Institut Telecom ex GET Groupe des Écoles des Télécommunications]

PhD Students

Maxime Anvari [FX-Conseil, PhD Student, from Nov 2019]

Lucas Benmouffok [Institut de recherche technologique System X, PhD Student]
Hanna-Mae Bisserier [Institut de recherche technologique System X, PhD Student]
Sarah Bordage [École polytechnique, PhD Student]
Alexis Challande [Quarkslab, PhD Student, from May 2019]
Mathilde Chenu de La Morinerie [École polytechnique, PhD Student]
Antonin Leroux [Ministère des armées, PhD Student, from Oct 2019]
Simon Montoya [Idemia, PhD Student, from Sep 2019]
Isabella Panaccione [Inria, PhD Student, from Oct 2018]
Maxime Roméas [École polytechnique, PhD Student, from Oct 2019]
Angelo Saadeh [Telecom ParisTech, PhD Student, from Sep 2019]

Post-Doctoral Fellows

Adrien Hauteville [Inria, Post-Doctoral Fellow, from Feb 2019]
Jade Nardi [Inria, Post-Doctoral Fellow, from Oct 2019]

Visiting Scientists

Gianira Alfarano [University of Zurich, from Sep 2019]
Vincent Neiger [Univ de Limoges, Nov 2019]
Sven Puchinger [DTY Lyngby, from Dec 2019]
Kazuhiro Yokoyama [Rikkyo University, Oct 2019]
Alessandro Neri [University of Zurich, from Sep 2019]

Administrative Assistant

Maria Agustina Ronco [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Scientific foundations

Grace combines expertise and deep knowledge in algorithmic number theory and algebraic geometry, to build and analyse (public-key) cryptosystems, design new error correcting codes, with real-world concerns like cybersecurity or blockchains (software and hardware implemntations, secure implementations in constrained environments, countermeasures against side channel attacks, white box cryptography).

The foundations of Grace therefore lie in algorithmic number theory (fundamental algorithms primality, factorization), number fields, the arithmetic geometry of curves, algebraic gemoetry and the theory of algebraic codes.

Arithmetic Geometry is the meeting point of algebraic geometry and number theory: the study of geometric objects defined over arithmetic number systems. In our case, the most important objects are curves and their Jacobians over finite fields; these are fundamental to our applications in both coding theory and cryptology. Jacobians of curves are excellent candidates for cryptographic groups when constructing efficient instances of public-key cryptosystems, of which Diffie–Hellman key exchange is an instructive example.

Coding Theory studies originated with the idea of using redundancy in messages to protect them against noise and errors. While the last decade of the 20th century has seen the success of so-called iterative decoding methods, we see now many new ideas in the realm of algebraic coding, with the foremost example being list decoding, (zero knowledge or not) proofs of computation.

Part of the activities of the team are oriented towards post-quantum cryptography, either based on elliptic curves (isogenies) or code-based. Also the team study relevant cryptography for the blockchain arena.

The group is strongly invested in cybersecurity: software security, secure hardware implementations, privacy, etc.

3. Research Program

3.1. Algorithmic Number Theory

Participants: Luca de Feo, François Morain, Benjamin Smith, Mathilde de La Morinerie, Antonin Leroux, Guénaël Renault.

Algorithmic Number Theory is concerned with replacing special cases with general algorithms to solve problems in number theory. In the Grace project, it appears in three main threads:

- fundamental algorithms for integers and polynomials (including primality and factorization);
- algorithms for finite fields (including discrete logarithms);
- algorithms for algebraic curves.

Clearly, we use computer algebra in many ways. Research in cryptology has motivated a renewed interest in Algorithmic Number Theory in recent decades—but the fundamental problems still exist *per se*. Indeed, while algorithmic number theory application in cryptanalysis is epitomized by applying factorization to breaking RSA public key, many other problems, are relevant to various area of computer science. Roughly speaking, the problems of the cryptological world are of bounded size, whereas Algorithmic Number Theory is also concerned with asymptotic results.

3.2. Arithmetic Geometry: Curves and their Jacobians

Participants: Luca de Feo, François Morain, Benjamin Smith, Mathilde de La Morinerie, Antonin Leroux.

Theme: Arithmetic Geometry: Curves and their Jacobians *Arithmetic Geometry* is the meeting point of algebraic geometry and number theory: that is, the study of geometric objects defined over arithmetic number systems (such as the integers and finite fields). The fundamental objects for our applications in both coding theory and cryptology are curves and their Jacobians over finite fields.

An algebraic *plane curve*$\mathcal{X}$ over a field $\mathbf{K}$ is defined by an equation

$$\mathcal{X} : F_{\mathcal{X}}(x, y) = 0 \quad \text{where} \ F_{\mathcal{X}} \in \mathbf{K}[x, y].$$

(Not every curve is planar—we may have more variables, and more defining equations—but from an algorithmic point of view, we can always reduce to the plane setting.) The *genus*$g_{\mathcal{X}}$ of $\mathcal{X}$ is a non-negative integer classifying the essential geometric complexity of $\mathcal{X}$; it depends on the degree of $F_{\mathcal{X}}$ and on the number of singularities of $\mathcal{X}$. The curve $\mathcal{X}$ is associated in a functorial way with an algebraic group $J_{\mathcal{X}}$, called the *Jacobian* of $\mathcal{X}$. The group $J_{\mathcal{X}}$ has a geometric structure: its elements correspond to points on a $g_{\mathcal{X}}$-dimensional projective algebraic group variety. Typically, we do not compute with the equations defining this projective variety: there are too many of them, in too many variables, for this to be convenient. Instead, we use fast algorithms based on the representation in terms of classes of formal sums of points on $\mathcal{X}$.

The simplest curves with nontrivial Jacobians are curves of genus 1, known as *elliptic curves*; they are typically defined by equations of the form $y^2 = x^3 + Ax + B$. Elliptic curves are particularly important given their central role in public-key cryptography over the past two decades. Curves of higher genus are important in both cryptography and coding theory.

3.3. Curve-Based cryptology

Participants: Luca de Feo, François Morain, Benjamin Smith, Mathilde de La Morinerie, Antonin Leroux.

Theme: Curve-Based Cryptology

Jacobians of curves are excellent candidates for cryptographic groups when constructing efficient instances of public-key cryptosystems. Diffie–Hellman key exchange is an instructive example.

Suppose Alice and Bob want to establish a secure communication channel. Essentially, this means establishing a common secret *key*, which they will then use for encryption and decryption. Some decades ago, they would have exchanged this key in person, or through some trusted intermediary; in the modern, networked world, this is typically impossible, and in any case completely unscalable. Alice and Bob may be anonymous parties who want to do e-business, for example, in which case they cannot securely meet, and they have no way to be sure of each other's identities. Diffie–Hellman key exchange solves this problem. First, Alice and Bob publicly agree on a cryptographic group G with a generator P (of order N); then Alice secretly chooses an integer a from $[1..N]$, and sends aP to Bob. In the meantime, Bob secretly chooses an integer b from $[1..N]$, and sends bP to Alice. Alice then computes $a(bP)$, while Bob computes $b(aP)$; both have now computed abP, which becomes their shared secret key. The security of this key depends on the difficulty of computing abP given P, aP, and bP; this is the Computational Diffie–Hellman Problem (CDHP). In practice, the CDHP corresponds to the Discrete Logarithm Problem (DLP), which is to determine a given P and aP.

This simple protocol has been in use, with only minor modifications, since the 1970s. The challenge is to create examples of groups G with a relatively compact representation and an efficiently computable group law, and such that the DLP in G is hard (ideally approaching the exponential difficulty of the DLP in an abstract group). The Pohlig–Hellman reduction shows that the DLP in G is essentially only as hard as the DLP in its largest prime-order subgroup. We therefore look for compact and efficient groups of prime order.

The classic example of a group suitable for the Diffie–Hellman protocol is the multiplicative group of a finite field $\mathbf{F}_q$. There are two problems that render its usage somewhat less than ideal. First, it has too much structure: we have a subexponential Index Calculus attack on the DLP in this group, so while it is very hard, the DLP falls a long way short of the exponential difficulty of the DLP in an abstract group. Second, there is only one such group for each q: its subgroup treillis depends only on the factorization of $q - 1$, and requiring $q - 1$ to have a large prime factor eliminates many convenient choices of q.

This is where Jacobians of algebraic curves come into their own. First, elliptic curves and Jacobians of genus 2 curves do not have a subexponential index calculus algorithm: in particular, from the point of view of the DLP, a generic elliptic curve is currently *as strong as* a generic group of the same size. Second, they provide some diversity: we have many degrees of freedom in choosing curves over a fixed $\mathbf{F}_q$, with a consequent diversity of possible cryptographic group orders. Furthermore, an attack which leaves one curve vulnerable may not necessarily apply to other curves. Third, viewing a Jacobian as a geometric object rather than a pure group allows us to take advantage of a number of special features of Jacobians. These features include efficiently computable pairings, geometric transformations for optimised group laws, and the availability of efficiently computable non-integer endomorphisms for accelerated encryption and decryption.

3.4. Algebraic Coding Theory

Participants: Daniel Augot, Alain Couvreur, Françoise Levy-Dit-Vehel, Maxime Roméas, Sarah Bordage, Adrien Hauteville, Isabella Panaccione.

Theme: Coding theory

Coding Theory studies originated with the idea of using redundancy in messages to protect against noise and errors. The last decade of the 20th century has seen the success of so-called iterative decoding methods, which enable us to get very close to the Shannon capacity. The capacity of a given channel is the best achievable transmission rate for reliable transmission. The consensus in the community is that this capacity is more easily reached with these iterative and probabilistic methods than with algebraic codes (such as Reed–Solomon codes).

However, algebraic coding is useful in settings other than the Shannon context. Indeed, the Shannon setting is a random case setting, and promises only a vanishing error probability. In contrast, the algebraic Hamming approach is a worst case approach: under combinatorial restrictions on the noise, the noise can be adversarial, with strictly zero errors.

These considerations are renewed by the topic of list decoding after the breakthrough of Guruswami and Sudan at the end of the nineties. List decoding relaxes the uniqueness requirement of decoding, allowing a small list of candidates to be returned instead of a single codeword. List decoding can reach a capacity close to the Shannon capacity, with zero failure, with small lists, in the adversarial case. The method of Guruswami and Sudan enabled list decoding of most of the main algebraic codes: Reed–Solomon codes and Algebraic–Geometry (AG) codes and new related constructions "capacity-achieving list decodable codes". These results open the way to applications again adversarial channels, which correspond to worst case settings in the classical computer science language.

Another avenue of our studies is AG codes over various geometric objects. Although Reed–Solomon codes are the best possible codes for a given alphabet, they are very limited in their length, which cannot exceed the size of the alphabet. AG codes circumvent this limitation, using the theory of algebraic curves over finite fields to construct long codes over a fixed alphabet. The striking result of Tsfasman–Vladut–Zink showed that codes better than random codes can be built this way, for medium to large alphabets. Disregarding the asymptotic aspects and considering only finite length, AG codes can be used either for longer codes with the same alphabet, or for codes with the same length with a smaller alphabet (and thus faster underlying arithmetic).

From a broader point of view, wherever Reed–Solomon codes are used, we can substitute AG codes with some benefits: either beating random constructions, or beating Reed–Solomon codes which are of bounded length for a given alphabet.

Another area of Algebraic Coding Theory with which we are more recently concerned is the one of Locally Decodable Codes. After having been first theoretically introduced, those codes now begin to find practical applications, most notably in cloud-based remote storage systems.

4. Application Domains

4.1. Application Domain: cybersecurity

Participants: Guénaël Renault, Benjamin Smith, François Morain, Alexis Challande, Simon Montoya, Maxime Anvari.

We are interesting in developing some interactions between cryptography and cybersecurity. In particular, we develop some researches in embedded security (side channels and fault attack), software security (finding vulnerability efficiently) and privacy (security of TOR).

4.2. Application Domain: blockchains

Participants: Daniel Augot, Sarah Bordage, Matthieu Rambaud, Lucas Benmouffok, Hanna-Mae Bisserier.

The huge interest shown by companies for blockchains and cryptocurrencies have attracted the attention of mainstream industries for new, advanced uses of cryptographic, beyond confidentiality, integrity and authentication. In particular, zero-knowledge proofs, computation with encrypted data, etc, are now revealing their potential in the blockchain context. Team Grace is investigating two topics in these areas: secure multiparty computation and so-called "STARKS".

Secure multiparty computation enables several participants to compute a common function of data they each secretly own, without each participant revealing his data to the other participants. This area has seen great progress in recent years, and the cryptogaphic protocols are now mature enough for practical use. This topic is new to project-team Grace, and we will investigate it in the context of blockchains, through the lenses of use for private "smart contracts". A PhD student has been hired since October, funded by IRT System-X.

Daniel Augot is involved in blockchains from the point of view of cryptography for better blockchains, mainly for improving privacy. A PhD student has been enrolled at IRT System-X, to study pratical use cases of Secure Multiparty Computaiton.

Also Daniel Augot, together with Julian Prat (economist, ENSAE), is leading a Polytechnique teaching and research “chair”, funded by CapGemini, for blockchains in the industry, B2B platforms, supply chains, etc.

4.3. Cloud storage

The team is concerned with several aspect of reliability and security of cloud storage, obtained mainly with tools from coding theory. On the privacy side, we build protocols for so-called Private Information Retrieval which enable a user to query a remote database for an entry, while not revealing his query. For instance, a user could query a service for stock quotes without revealing with company he is interested in. On the availability side, we study protocols for proofs of retrievability, which enable a user to get assurance that a huge file is still available on a remote server, with a low bandwith protocol which does not require to download the whole file. For instance, in a peer-to-peer distributed storage system, where nodes could be rewarded for storing data, they can be audited with proof of retrievability protocols to make sure they indeed hold the data.

We investigate these problems with algebraic coding theory, mainly codes with locality (locally decodable codes, locally recoverable codes, and so on).

An M2 intern, Maxime Roméas, Bordeaux university, studied the constructive cryptography model, "A study of the Constructive Cryptography model of Maurer et. al." 5 months, followed by a PhD grant from IP Paris/Ecole Polytechnique for a 3-year doctorate (Oct 2019-Sept 2022): "The Constructive Cryptography paradigm applied to Interactive Cryptographic Proofs".

The Constructive Cryptography framework redefines basic cryptographic primitives and protocols starting from discrete systems of three types (resources, converters, and distinguishers). This not only permits to construct them effectively, but also lighten and sharpen their security proofs. One strength of this model is its composability. The purpose of the PhD is to apply this model to rephrase existing interactive cryptographic proofs so as to assert their genuine security, as well as to design new proofs. The main concern here is security and privacy in Distributed Storage settings.

5. New Software and Platforms

5.1. ACTIS

Algorithmic Coding Theory in Sage

FUNCTIONAL DESCRIPTION: The aim of this project is to vastly improve the state of the error correcting library in Sage. The existing library does not present a good and usable API, and the provided algorithms are very basic, irrelevant, and outdated. We thus have two directions for improvement: renewing the APIs to make them actually usable by researchers, and incorporating efficient programs for decoding, like J. Nielsen’s CodingLib, which contains many new algorithms.

- Partner: Technical University Denmark
- Contact: Daniel Augot

5.2. DECODING

KEYWORD: Algebraic decoding

FUNCTIONAL DESCRIPTION: Decoding is a standalone C library. Its primary goal is to implement Guruswami–Sudan list decoding-related algorithms, as efficiently as possible. Its secondary goal is to give an efficient tool for the implementation of decoding algorithms (not necessarily list decoding algorithms) and their benchmarking.

- Participant: Guillaume Quintin
- Contact: Daniel Augot

5.3. Fast Compact Diffie-Hellman

KEYWORD: Cryptography

FUNCTIONAL DESCRIPTION: A competitive, high-speed, open implementation of the Diffie–Hellman protocol, targeting the 128-bit security level on Intel platforms. This download contains Magma files that demonstrate how to compute scalar multiplications on the x-line of an elliptic curve using endomorphisms. This accompanies the EuroCrypt 2014 paper by Costello, Hisil and Smith, the full version of which can be found here: http://eprint.iacr.org/2013/692. The corresponding SUPERCOP-compatible crypto_dh application can be downloaded from http://hhisil.yasar.edu.tr/files/hisil20140318compact.tar.gz.

- Participant: Ben Smith
- Contact: Ben Smith
- URL: http://research.microsoft.com/en-us/downloads/ef32422a-af38-4c83-a033-a7aafbc1db55/

5.4. CADO-NFS

Crible Algébrique: Distribution, Optimisation - Number Field Sieve

KEYWORDS: Cryptography - Number theory

FUNCTIONAL DESCRIPTION: CADO-NFS is a complete implementation in C/C++ of the Number Field Sieve (NFS) algorithm for factoring integers and computing discrete logarithms in finite fields. It consists in various programs corresponding to all the phases of the algorithm, and a general script that runs them, possibly in parallel over a network of computers.

NEWS OF THE YEAR: The main program for relation collection now supports composite "special-q". The memory footprint of the central step of linear algebra was reduced. Parallelism of many of the Cado-NFS programs was improved considerably (sieving, relation filtering, as well as the central step of linear algebra).

- Participants: Pierrick Gaudry, Emmanuel Thomé and Paul Zimmermann
- Contact: Emmanuel Thomé
- URL: http://cado-nfs.gforge.inria.fr/

6. New Results

6.1. Error Locating pairs

Participants: Alain Couvreur, Isabella Panaccione.

Algebraic codes such as Reed–Solomon codes and algebraic geometry codes benefit from efficient decoding algorithms permitting to correct errors up to half the minimum distance and sometimes beyond. In 1992, Pellikaan proved that many **unique** decoding could be unified using an object called *Error correcting pair*. In short, given an error correcting code $\mathcal{C}$, an error correcting pair for $\mathcal{C}$ is a pair of codes $(\mathcal{A}, \mathcal{B})$ whose component wise product $\mathcal{A} * \mathcal{B}$ is contained in the dual code $\mathcal{C}^{\perp}$ and such that $\mathcal{A}, \mathcal{B}$ satisfy some constraints of dimension and minimum distance.

On the other hand, in the late 90's, after the breakthrough of Sudan and Guruswami Sudan the question of list decoding permitting to decode beyond half the minimum distance. In a recently submitted article, A. Couvreur and I. Panaccione [15] proposed a unified point of view for probabilistic decoding algorithms decoding beyond half the minimum distance. Similarly to Pellikaan's result, this framework applies to any code benefiting from an *error locating pair* which is a relaxed version of error correcting pairs.

6.2. Factoring oracles

Participants: François Morain, Benjamin Smith, Guénaël Renault.

Integer factoring is an old topic, and the situation is as follows: in the classical world, we think integer factoring is hard and the algorithms we have are quite powerful though of subexponential complexity and factoring numbers with several hundred bits; whereas in the quantum world, it is assumed to be easy (i.e., there exists a quantum polynomial time algorithm) but never experienced and the record is something like a few bits. F. Morain, helped by B. Smith and G. Renault studied the theoretical problem of factoring integers given access to classical oracles, like the Euler totient function. They were able to give some interesting classes of numbers that could tackled, The manuscript [18] is currently being refereed.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Participants: Daniel Augot, Alain Couvreur, Guénaël Renault, François Morain.

- Through École polytechnique, Daniel Augot is leader of a teaching and research chair on Blockchains for business, funded by CapGemini.
- IRT System-X funds a PhD student for Secure Multiparty Computation in blockchains
- Ernst & Young funds a contract for providing PhD guidance to one of its employee, on the topic of blockchains
- Idemia funds a CIFRE PhD student on the secure implementation in constrained environement of post-quantum cryptosystems.
- Quarkslab funds a CIFRE PhD student on the analysis of malware code
- French Min. Arm. funds a PhD student on the analysis of the ToR network
- Grant with Nokia with the Privacy “Action de recherche”.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Participants: Daniel Augot, Matthieu Rambaud.

Daniel Augot and Matthieu Rambaud (Institut Mines-Telecom) received a Digicosme Grant, to fund a new PhD student, A. Saadeh, starting November 2019, on the topic of Secure Multiparty Computation.

8.2. National Initiatives

8.2.1. ANR MANTA

Participants: Daniel Augot, Alain Couvreur, Françoise Levy-Dit-Vehel, Philippe Lebacque, Matthieu Rambaud, Isabella Panaccione, Luca de Feo.

MANTA (accepted July 2015, starting March 2016, Ended September 2019): “Curves, surfaces, codes and cryptography”. This project deals with applications of coding theory error correcting codes to in cryptography, multi-party computation, and complexity theory, using advanced topics in algebraic geometry and number theory.

We have four annual national retreats, the last one in January 2019, and we organized a closing international workshop in August 2019, with more than 40 participants, half French, half international.

See http://anr-manta.inria.fr/.

8.2.2. ANR CIAO

Participants: Benjamin Smith, Luca de Feo, Antonin Leroux, Mathilde de La Morinerie.

ANR **CIAO** (Cryptography, Isogenies, and Abelian varieties Overwhelming) is a JCJC 2019 project, led by Damien Robert (Inria EP LFANT). This project, which started in October 2019, will examine applications of higher-dimensional abelian varieties in isogeny-based cryptography.

8.2.3. ANR CBCRYPT

Participant: Alain Couvreur.

ANR **CBCRYPT** (Code–based Cryptography) This is a project from (*Appel à projets générique, Défi 9, Liberté et sécurité de l'Europe, de ses citoyens et de ses résidents, Axe 4 ; Cybersécurité*). This project, starting in october 2017 led by Jean-Pierre Tillich (Inria, EP Cosmiq) focusses on the design and the security analysis of code–based primitives, in the context of the current NIST competition.

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

Participant: Benjamin Smith.

- SPARTA https://www.sparta.eu/ is a cybersecurity competence network, with the objective to collaboratively develop and implement top-tier research and innovation actions

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Alessandro Neri visited us from September 2019 to December 2019, as post-doctoral visitor, to work on rank-metric codes.
- Vincent Neiger (Mcf, Univ. Limoges) visited our team twice. One week in march and one meek in november, to work on the decoding of Reed–Solomon codes.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

9.1.1.1. Member of the Conference Program Committees

- Tokenomics 2019, International Conference on Blockchain Economics, Security and Protocols, Paris: D. Augot.
- FAB 2019, Second International Symposium on Foundations and Applications of Blockchain, Los Angeles: D. Augot
- ICBC 2019 (IEEE International Conference on Blockchain and Cryptocurrency, Seoul): D. Augot
- CBT 2019 (3rd International Workshop on Cryptocurrencies and Blockchain Technology, Barcelona)
- ECC 2019 (23rd International Workshop on Elliptic Curve Cryptography, Bochum): B. Smith
- Latincrypt 2019 (Santiago de Chile): B. Smith
- C2SI (*Codes, Cryptographie et Sécurité Informatique*) 2019 (Rabat, Morocco) A. Couvreur

9.1.1.2. Reviewer

- Eurocrypt 2019: D. Augot, B. Smith

- Indocrypt 2019 (20th International Conference on Cryptology in India, Hyderabad):D. Augot
- ISIT (International Symposium on Information Theory) 2019: D. Augot, A. Couvreur.
- Latincrypt 2019. A. Couvreur.
- SAC 2019: B. Smith
- STACS 2020: B. Smith

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- F. Morain is member of the editorial board of the *Applicable Algebra in Engineering, Communication and Computing*, Springer.
- A. Couvreur is member of the editorial board of the *Publications mathématiques de l'Institut de mathématiques de Besançon, Algèbre et Théorie des nombres*.

9.1.2.2. Reviewer - Reviewing Activities

- Applicable Algebra in Engineering, Communication, and Computing: B. Smith
- Journal of Cryptographic Engineering: B. Smith
- Journal of Cryptology: B. Smith
- Publications Mathématiques de Besançon: B. Smith
- Transactions on Mathematical Software: B. Smith
- Designs, Codes and Cryptography: A. Couvreur.
- IEEE, Transactions on Information Theory: A. Couvreur.
- IEEE, Transactions on Communication: A. Couvreur.

9.1.3. Invited Talks

- D. Augot was invited to the joint Caen-Rouen ArcoCrypt colloquium.
- F. Morain was invited to give a talk at the seminar of the ARIC project-team in Lyon.
- G. Renault was invited to give the main keynote at PHISIC'19 workshop (Gardanne)
- G. Renault was invited to give a talk at the Workshop on Randomness and Arithmetics for Cryptography on Hardware (Roscoff)
- G. Renault was invited to give a talk at the Journée Internationale Post-Quantique organized by Institut Cyber de Grenoble Alpes.
- B. Smith was invited to give a talk at the Workshop on Arithmetic of low-dimensional abelian varieties at ICERM (Providence, USA)
- B. Smith and A. Couvreur were invited to give a talk in the mini-symposium on isogeny-based cryptography at the SIAM Conference on Applied Algebraic Geometry (Bern, Switzerland)
- B. Smith was invited to give a talk in the Autumn session of *Arithmétique en Plat Pays* (Mons, Belgium)
- A. Couvreur was invited speaker at the conference *WCC (Workshop on Coding and Cryptography) 2019* Saint Jacut de la mer, France.
- A. Couvreur was invited speaker at the conference *NuTMIC (Number Theoretic Methods In Cryptography) 2019*, Paris.

9.1.4. Seminars

D. Augot is member of the scientific committee of the C2 seminar, which is the French wide, now itinerant, seminar of the subgroup "Codage et Cryptographie" of the CNRS GDR group "Informatique mathématique".

9.1.5. Scientific Expertise

G. Renault was member of the Comité d'Évaluation du LJK (Grenoble) pour l'Hcéres.

9.1.6. Research Administration

- F. Morain is vice-head of the Département d'informatique of Ecole Polytechnique; in charge of years 1 and 2 for Computer Science courses.
- F. Morain is member of the Board of Master Parisien de Recherche en Informatique (MPRI); also a member of the board of the Cybersecurity track in the CS Master of IPParis.
- Recruitment committees:
  - D. Augot participated in a selection committee for an Assistant Professor position at École polytechnique.
  - A. Couvreur is member of the *commission scientifique* of Inria Saclay's research centre.
- Funding
  - D. Augot belongs to the Inria-NomadicLabs committee.
  - D. Augot belongs to MATH-INFO subcommittee of Saclay labex Laboratoire Jacques Hadamard, and has been replaced by A. Couvreur.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : F. Morain, Lectures for INF361: "Introduction à l'informatique", 15h (equiv TD), 1st year (L3), École polytechnique. Coordinator of this module (350 students).

Licence : B. Smith: *CSE101: Introduction to Computer Programming*, 42h, L1, École polytechnique, France

Licence : G. Renault: INF361, *Introduction à l'informatique*, 50h, L3, École Polytechnique, France.

Master : A. Couvreur : *MPRI 2-13-2: Error Correcting codes and applications to cryptography*.

Master : D. Augot: lectures and labs on crypto in blockchains, 24h, M2, École polytechnique, France.

Master : F. Morain is the scientific leader of the Master of Science and Technology *Cybersecurity: Threats and Defense* of École Polytechnique.

Master : F. Morain and A. Couvreur, INF558, *Introduction to cryptology*, 36h, M1, École Polytechnique.

Master : B. Smith: *INF568: Advanced Cryptography*, 54h, M1, École polytechnique, France

Master : B. Smith and F. Morain: *MPRI 2-12-2: Algorithmes Arithmétiques pour la Cryptologie*, 22.5h, M2, Master Parisien de Recherche en Informatique, France

Master : F. Levy-dit-Vehel, discrete maths, 21h, M1, ENSTA.

Master : F. Levy-dit-Vehel, cryptography, 24h, M2, ENSTA.

9.2.2. Supervision

HdR : A. Couvreur, Codes algébriques et géométriques, applications à la cryptographie et à l'information quantique, Paris Diderot University, December 16, 2019.

9.2.3. Juries

- D. Augot was member of the thesis committee of Thomas Debris.
- G. Renault was president of the thesis committee of François Boutigny.
- G. Renault was member of the thesis committee of Ramtine Tofighi.

- B. Smith was a member of the thesis committee of Louiza Papachristodoulou (Radboud Universiteit Nijmegen)
- B. Smith was a member of the thesis committee of Joost Renes (Radboud Universiteit Nijmegen)
- B. Smith was a *rapporteur* on the thesis of Yan Bo Ti (University of Auckland).

9.3. Popularization

9.3.1. Interventions...

- D. Augot was invited to the mathematical colloquium of the University of Besançon.
- Colloque "Blockchains et compétences" à l'Assemblée nationale le 14 mars: D. Augot, who participated to three meetings at Ministry of Finance and Ministry of Industry about blockchains.
- A. Couvreur is *Correspondant de Médiation Scientifique* of Inria Saclay's research centre.
- A. Couvreur organised the *Fête de la science 2019* at Inria Saclay on october 10 and 11 2019. J. Nardi, S. Bordage, M. Chenu de la Morinerie, M. Romeas participated to the event as volunteers.
- A. Couvreur organised the Rendez-vous des Jeunes Mathématiciennes et Informaticiennes (RJMI) at Inria Saclay on october 21 and 22, 2019.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] A. COUVREUR.*Codes algébriques et géométriques, applications à la cryptographie et à l'information quantique*, Université Paris Diderot, December 2019, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-02438668

Articles in International Peer-Reviewed Journal

[2] B. AUDOUX, A. COUVREUR.*On tensor products of CSS Codes*, in "Annales de l'Institut Henri Poincaré (D) Combinatorics, Physics and their Interactions", 2019, vol. 6, n^{o} 2, p. 239–287, https://arxiv.org/abs/1512.07081 [*DOI :* 10.4171/AIHPD/71], https://hal.archives-ouvertes.fr/hal-01248760

[3] N. COXON.*Fast Hermite interpolation and evaluation over finite fields of characteristic two*, in "Journal of Symbolic Computation", July 2019, https://arxiv.org/abs/1807.00645 [*DOI :* 10.1016/J.JSC.2019.07.014], https://hal.archives-ouvertes.fr/hal-01827583

[4] J. LAVAUZELLE.*Private Information Retrieval from Transversal Designs*, in "IEEE Transactions on Information Theory", 2019, vol. 65, n^{o} 2, p. 1189-1205, https://arxiv.org/abs/1709.07952 [*DOI :* 10.1109/TIT.2018.2861747], https://hal.archives-ouvertes.fr/hal-01901014

[5] J. LAVAUZELLE, F. LEVY-DIT-VEHEL.*Generic constructions of PoRs from codes and instantiations*, in "Journal of Mathematical Cryptology", February 2019, vol. 13, n^{o} 2, p. 81–106, forthcoming [*DOI :* 10.1515/JMC-2018-0018], https://hal.archives-ouvertes.fr/hal-02053948

[6] R. LERCIER, C. RITZENTHALER, F. ROVETTA, J. SIJSLING, B. SMITH.*Distributions of traces of Frobenius for smooth plane curves over finite fields*, in "Experimental Mathematics", January 2019, vol. 28, n^{o} 1, p. 39-48, https://arxiv.org/abs/1510.05601 [*DOI :* 10.1080/10586458.2017.1328321], https://hal.inria.fr/hal-01217995

International Conferences with Proceedings

[7] M. Bardet, M. Bertin, A. Couvreur, A. Otmani. *Practical Algebraic Attack on DAGS*, in "CBC 2019 - 7th Code-Based Cryptography Workshop", Darmstadt, Germany, LNCS, July 2019, vol. 11666, p. 86-101, https://arxiv.org/abs/1905.03635 - 16 pages, accepted for publication in the 7th Code-Based Cryptography Workshop 2019 [*DOI :* 10.1007/978-3-030-25922-8_5], https://hal.archives-ouvertes.fr/hal-02125330

[8] D. Cervantes-Vázquez, M. Chenu, J.-J. Chi-Domínguez, L. De Feo, F. Rodríguez-Henríquez, B. Smith. *Stronger and Faster Side-Channel Protections for CSIDH*, in "Latincrypt 2019 - 6th International Conference on Cryptology and Information Security in Latin", Santiago de Chile, Chile, P. Schwabe, N. Thériault (editors), LNCS - Lecture Notes in Computer Science, Springer, October 2019, vol. 11774, https://arxiv.org/abs/1907.08704 - This work has been accepted in LATINCRYPT-2019 [*DOI :* 10.1007/978-3-030-30530-7_9], https://hal.inria.fr/hal-02190863

[9] A. Couvreur, M. Lequesne, J.-P. Tillich. *Recovering short secret keys of RLCE encryption scheme in polynomial time*, in "PQCrypto 2019 - International Conference on Post-Quantum Cryptography", Chongqing, China, May 2019, https://arxiv.org/abs/1805.11489 [*DOI :* 10.1007/978-3-030-25510-7_8], https://hal.inria.fr/hal-01959617

Conferences without Proceedings

[10] D. Augot, H. Chabanne, W. George. *Practical Solutions to Save Bitcoins Applied to an Identity System Proposal*, in "5th International Conference on Information Systems Security and Privacy", Prague, France, SCITEPRESS - Science and Technology Publications, February 2019, p. 511-518 [*DOI :* 10.5220/0007443905110518], https://hal.telecom-paristech.fr/hal-02347238

[11] D. Coggia, A. Couvreur. *On the security of a Loidreau's rank metric code based encryption scheme*, in "WCC 2019 - The Eleventh International Workshop on Coding and Cryptography", Saint Jacut de la mer, France, March 2019, https://hal.archives-ouvertes.fr/hal-02064465

Scientific Books (or Scientific Book chapters)

[12] D. Augot, F. Solov'Eva, M. Minier, V. A. Zinoviev, T. Johansson. *Editorial: Special issue on coding and cryptography*, Springer, March 2019, vol. 87, n^{o} 2-3 [*DOI :* 10.1007/s10623-018-00601-w], https://hal.inria.fr/hal-02373772

Other Publications

[13] W. Castryck, T. Decru, B. Smith. *Hash functions from superspecial genus-2 curves using Richelot isogenies*, June 2019, https://arxiv.org/abs/1903.06451 - working paper or preprint, https://hal.inria.fr/hal-02067885

[14] C. Costello, B. Smith. *The supersingular isogeny problem in genus 2 and beyond*, December 2019, https://arxiv.org/abs/1912.00701 - working paper or preprint, https://hal.inria.fr/hal-02389073

[15] A. Couvreur, I. Panaccione. *Power Error Locating Pairs*, July 2019, https://arxiv.org/abs/1907.11658 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02196650

[16] N. Coxon. *Fast transforms over finite fields of characteristic two*, September 2019, https://arxiv.org/abs/1807.07785 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01845238

[17] J. LAVAUZELLE, P. LOIDREAU, B.-D. PHAM.*RAMESSES, a Rank Metric Encryption Scheme with Short Keys*, January 2020, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02426624

References in notes

[18] F. MORAIN, G. RENAULT, B. SMITH.*Deterministic factoring with oracles*, February 2018, https://arxiv.org/abs/1802.08444 - working paper or preprint, https://hal.inria.fr/hal-01715832

Project-Team ILDA

Interacting with Large Data

IN COLLABORATION WITH: Laboratoire de recherche en informatique (LRI)

IN PARTNERSHIP WITH:
CNRS

Université Paris-Sud (Paris 11)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Interaction and visualization

Table of contents

Project-Team ILDA

Creation of the Team: 2015 January 01, updated into Project-Team: 2016 December 01

Keywords:

Computer Science and Digital Science:

A3.1.7. - Open data
A3.1.10. - Heterogeneous data
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.2.6. - Linked data
A5.1. - Human-Computer Interaction
A5.1.6. - Tangible interfaces
A5.2. - Data visualization
A5.5.4. - Animation
A5.6.1. - Virtual reality
A5.6.2. - Augmented reality

Other Research Topics and Application Domains:

B9.5.3. - Physics
B9.5.6. - Data science
B9.6.7. - Geography
B9.7.2. - Open data
B9.11. - Risk management

1. Team, Visitors, External Collaborators

Research Scientists

Emmanuel Pietriga [Team leader, Inria, Senior Researcher, HDR]
Caroline Appert [CNRS, Senior Researcher, HDR]
Olivier Chapuis [CNRS, Researcher]

Faculty Member

Anastasia Bezerianos [Univ Paris-Sud, Associate Professor]

Technical Staff

Léo Colombaro [Inria, Engineer, until May 2019]
Adhitya Kamakshidasan [Inria, Engineer, until Sep 2019]
Dylan Lebout [Inria, Engineer]

PhD Students

Eugénie Brasier [Univ Paris-Sud, PhD Student]
Emmanuel Courtoux [Univ Paris-Sud, PhD Student, from Oct 2019]
Marie Destandau [Inria, PhD Student]
Anna Gogolou [Inria, PhD Student, until Nov 2019]
Raphaël James [Univ Paris-Sud, PhD Student]
Hugo Romat [Tecknowmetrix, PhD Student, until Oct 2019, granted by CIFRE]
Tong Xue [Inria, PhD Student]

Post-Doctoral Fellow

Vanessa Peña-Araya [Inria, Post-Doctoral Fellow]

Administrative Assistants

Alexandra Merlin [Inria, Administrative Assistant, until Sep 2019]
Katia Evrat [Inria, Administrative Assistant, since Nov 2019]

2. Overall Objectives

2.1. Overall Objectives

In an increasing number of domains, computer users are faced with large datasets, that are often interlinked and organized according to elaborate structures thanks to new data models such as those that are arising with the development of, *e.g.*, the Web of Data. Rather than seeing the inherent complexity of those data models as a hindrance, we aim at leveraging it to design new interactive systems that can better assist users in their data understanding and processing tasks.

These "Data-centric Interactive Systems" aim at providing users with the right information at the right time, presenting it in the most meaningful manner, and letting users efficiently manipulate, edit and share these data with others. This entails minimizing the effort required to retrieve and relate data from relevant sources; displaying data using visual presentation techniques that match the data's characteristics and the users' tasks; and providing users with means of interacting with the data that effectively support their train of thought.

Our approach is based on the idea of bringing the fields of Web data management [27] and Human-computer interaction [54], [79] closer together, based on the strong belief that they have the potential to cross-fertilize one another. User interface design is essential to the management and understanding of large, interlinked datasets. Interlinked datasets enriched with even a small amount of semantics have the potential to help create interfaces that let users analyze and manipulate data in a more efficient manner by providing them with, *e.g.*, more relevant query results and giving them efficient means to navigate and relate those results. Our ultimate, long-term goal is to design interactive systems that make it as straightforward to manipulate large webs of data as spreadsheets do for tabular data.

3. Research Program

3.1. Introduction

Our ability to acquire or generate, store, process, interlink and query data has increased spectacularly over the last few years. The corresponding advances are commonly grouped under the umbrella of so called *Big Data*. Even if the latter has become a buzzword, these advances are real, and they are having a profound impact in domains as varied as scientific research, commerce, social media, industrial processes or e-government. Yet, looking ahead, emerging technologies related to what we now call the *Web of Data* (a.k.a the Semantic Web) have the potential to create an even larger revolution in data-driven activities, by making information accessible to machines as semistructured data [26] that eventually becomes actionable knowledge. Indeed, novel Web data models considerably ease the interlinking of semi-structured data originating from multiple independent sources. They make it possible to associate machine-processable semantics with the data. This in turn means that heterogeneous systems can exchange data, infer new data using reasoning engines, and that software agents can cross data sources, resolving ambiguities and conflicts between them [77]. Datasets are becoming very rich and very large. They are gradually being made even larger and more heterogeneous, but also much more useful, by interlinking them, as exemplified by the Linked Data initiative [49].

These advances raise research questions and technological challenges that span numerous fields of computer science research: databases, communication networks, security and trust, data mining, as well as human-computer interaction. Our research is based on the conviction that interactive systems play a central role in many data-driven activity domains. Indeed, no matter how elaborate the data acquisition, processing and storage pipelines are, data eventually get processed or consumed one way or another by users. The latter are faced with large, increasingly interlinked heterogeneous datasets (see, *e.g.*, Figure 1) that are organized according to complex structures, resulting in overwhelming amounts of both raw data and structured information. Users thus require effective tools to make sense of their data and manipulate them.

Figure 1. Linking Open Data cloud diagram from 2007 to 2017 – http://lod-cloud.net

We approach this problem from the perspective of the Human-Computer Interaction (HCI) field of research, whose goal is to study how humans interact with computers and inspire novel hardware and software designs aimed at optimizing properties such as efficiency, ease of use and learnability, in single-user or cooperative work contexts. More formally, HCI is about designing systems that lower the barrier between users' cognitive model of what they want to accomplish, and computers' understanding of this model. HCI is about the design, implementation and evaluation of computing systems that humans interact with [54], [79]. It is a highly multidisciplinary field, with experts from computer science, cognitive psychology, design, engineering, ethnography, human factors and sociology.

In this broad context, ILDA aims at designing interactive systems that display [35], [61], [87] the data and let users interact with them, aiming to help users better *navigate* and *comprehend* large webs of data represented visually [7], as well as *relate* and *manipulate* them.

Our research agenda consists of the three complementary axes detailed in the following subsections. Designing systems that consider interaction in close conjunction with data semantics is pivotal to all three axes. Those semantics will help drive navigation in, and manipulation of, the data, so as to optimize the communication bandwidth between users and data.

3.2. Semantics-driven Data Manipulation

Participants: Emmanuel Pietriga, Caroline Appert, Anastasia Bezerianos, Marie Destandau, Hugo Romat, Tong Xue, Léo Colombaro.

The Web of Data has been maturing for the last fifteen years and is starting to gain adoption across numerous application domains (Figure 1). Now that most foundational building blocks are in place, from knowledge representation, inference mechanisms and query languages [50], all the way up to the expression of data presentation knowledge [70] and to mechanisms like look-up services [86] or spreading activation [43], we need to pay significant attention to how human beings are going to interact with this new Web, if it is to *"reach its full potential"* [44].

Most efforts in terms of user interface design and development for the Web of data have essentially focused on tools for software developers or subject-matter experts who create ontologies and populate them [56], [41]. Tools more oriented towards end-users are starting to appear [32], [34], [51], [52], [55], [64], including the so-called *linked data browsers* [49]. However, those browsers are in most cases based on quite conventional point-and-click hypertext interfaces that present data to users in a very page-centric, web-of-documents manner that is ill-suited to navigating in, and manipulating, webs of data.

To be successful, interaction paradigms that let users navigate and manipulate data on the Web have to be tailored to the radically different way of browsing information enabled by it, where users directly interact with the data rather than with monolithic documents. The general research question addressed in this part of our research program is how to design novel interaction techniques that help users manipulate their data more efficiently. By data manipulation, we mean all low-level tasks related to manually creating new content, modifying and cleaning existing content, merging data from different sources, establishing connections between datasets, categorizing data, and eventually sharing the end results with other users; tasks that are currently considered quite tedious because of the sheer complexity of the concepts, data models and syntax, and the interplay between all of them.

Our approach is based on the conviction that there is a strong potential for cross-fertilization, as mentioned earlier: on the one hand, user interface design is essential to the management and understanding of webs of data; on the other hand, interlinked datasets enriched with even a small amount of semantics can help create more powerful user interfaces, that provide users with the right information at the right time.

We envision systems that focus on the data themselves, exploiting the underlying *semantics and structure* in the background rather than exposing them – which is what current user interfaces for the Web of Data often do. We envision interactive systems in which the semantics and structure are not exposed directly to users, but serve as input to the system to generate interactive representations that convey information relevant to the task at hand and best afford the possible manipulation actions.

Relevant publications by team members this year: [22], [15], [17] and major ones in recent years: [7].

3.3. Generalized Multi-scale Navigation

Participants: Caroline Appert, Anastasia Bezerianos, Olivier Chapuis, Emmanuel Pietriga, Vanessa Peña-Araya, Marie Destandau, Anna Gogolou, Hugo Romat, Dylan Lebout.

The foundational question addressed here is what to display when, where and how, so as to provide effective support to users in their data understanding and manipulation tasks. ILDA targets contexts in which workers have to interact with complementary views on the same data, or with views on different-but-related datasets, possibly at different levels of abstraction. Being able to combine or switch between representations of the data at different levels of detail and merge data from multiple sources in a single representation is central to many scenarios. This is especially true in both of the application domains we consider: mission-critical systems (e.g., natural disaster crisis management) and the exploratory analysis of scientific data (e.g., correlate theories and heterogeneous observational data for an analysis of a given celestial body in Astrophysics).

A significant part of our research over the last ten years has focused on multi-scale interfaces. We designed and evaluated novel interaction techniques, but also worked actively on the development of open-source UI toolkits for multi-scale interfaces (http://zvtm.sf.net). These interfaces let users navigate large but relatively homogeneous datasets at different levels of detail, on both workstations [73], [29], [69], [68], [67], [30], [72], [28], [74] and wall-sized displays [63], [58], [71], [62], [31], [37], [36]. This part of the ILDA research program is about extending multi-scale navigation in two directions: 1. Enabling the representation of multiple, spatially-registered but widely varying, multi-scale data layers in Geographical Information Systems (GIS); 2. Generalizing the multi-scale navigation paradigm to interconnected, heterogeneous datasets as found on the Web of Data.

The first research problem has been mainly investigated in collaboration with IGN in the context of ANR project MapMuxing, which stands for *multi-dimensional map multiplexing*, from 2014 to early 2019. Project MapMuxing aimed at going beyond the traditional pan & zoom and overview+detail interface schemes, and at designing and evaluating novel cartographic visualizations that rely on high-quality generalization, *i.e.*, the simplification of geographic data to make it legible at a given map scale [82], [83], and symbol specification. Beyond project MapMuxing, we are also investigating multi-scale multiplexing techniques for geo-localized data in the specific context of ultra-high-resolution wall-sized displays, where the combination of a very high pixel density and large physical surface enable us to explore designs that involve collaborative interaction and physical navigation in front of the workspace. This is work done in cooperation with team Massive Data at Inria Chile.

The second research problem is about the extension of multi-scale navigation to interconnected, heterogeneous datasets. Generalization has a rather straightforward definition in the specific domain of geographical information systems, where data items are geographical entities that naturally aggregate as scale increases. But it is unclear how generalization could work for representations of the more heterogeneous webs of data that we consider in the first axis of our research program. Those data form complex networks of resources with multiple and quite varied relationships between them, that cannot rely on a single, unified type of representation (a role played by maps in GIS applications).

Addressing the limits of current generalization processes is a longer-term, more exploratory endeavor. Here again, the machine-processable semantics and structure of the data give us an opportunity to rethink how users navigate interconnected heterogeneous datasets. Using these additional data, we investigate ways to generalize the multi-scale navigation paradigm to datasets whose layout and spatial relationships can be much richer and much more diverse than what can be encoded with static linear hierarchies as typically found today in interfaces for browsing maps or large imagery. Our goal is thus to design and develop highly dynamic and versatile multi-scale information spaces for heterogeneous data whose structure and semantics are not known in advance, but discovered incrementally.

Relevant publications by team members this year: [24], [20], [13], [14], [11], [19] and major ones in recent years: [10], [2].

3.4. Novel Forms of Input for Groups and Individuals

Participants: Caroline Appert, Anastasia Bezerianos, Olivier Chapuis, Emmanuel Pietriga, Eugénie Brasier, Emmanuel Courtoux, Raphaël James.

Analyzing and manipulating large datasets can involve multiple users working together in a coordinated manner in multi-display environments: workstations, handheld devices, wall-sized displays [31]. Those users work towards a common goal, navigating and manipulating data displayed on various hardware surfaces in a coordinated manner. Group awareness [48], [25] is central in these situations, as users, who may or may not be co-located in the same room, can have an optimal individual behavior only if they have a clear picture of what their collaborators have done and are currently doing in the global context. We work on the design and implementation of interactive systems that improve group awareness in co-located situations [57], making individual users able to figure out what other users are doing without breaking the flow of their own actions.

In addition, users need a rich interaction vocabulary to handle large, structured datasets in a flexible and powerful way, regardless of the context of work. Input devices such as mice and trackpads provide a limited number of input actions, thus requiring users to switch between modes to perform different types of data manipulation and navigation actions. The action semantics of these input devices are also often too much dependent on the display output. For instance, a mouse movement and click can only be interpreted according to the graphical controller (widget) above which it is moved. We focus on designing powerful input techniques based upon technologies such as tactile surfaces (supported by UI toolkits developed in-house), 3D motion tracking systems, or custom-built controllers [60] *to complement (rather than replace) traditional input devices* such as keyboards, that remain the best method so far for text entry, and indirect input devices such as mice or trackpads for pixel-precise pointing actions.

The input vocabularies we investigate enable users to navigate and manipulate large and structured datasets in environments that involve multiple users and displays that vary in their size, position and orientation [31], [45], each having their own characteristics and affordances: wall displays [63], [89], workstations, tabletops [66], [40], tablets [65], [84], smartphones [88], [38], [80], [81], and combinations thereof [39], [85], [62], [31].

We aim at designing rich interaction vocabularies that go far beyond what current touch interfaces offer, which rarely exceeds five gestures such as simple slides and pinches. Designing larger gesture vocabularies requires identifying discriminating dimensions (e.g., the presence or absence of anchor points and the distinction between internal and external frames of reference [65]) in order to structure a space of gestures that interface designers can use as a dictionary for choosing a coherent set of controls. These dimensions should be few and simple, so as to provide users with gestures that are easy to memorize and execute. Beyond gesture complexity, the scalability of vocabularies also depends on our ability to design robust gesture recognizers that will allow users to fluidly chain simple gestures that make it possible to interlace navigation and manipulation actions.

We also study how to further extend input vocabularies by combining touch [65], [88], [66] and mid-air gestures [63] with physical objects [53], [78], [60] and classical input devices such as keyboards to enable users to input commands to the system or to involve other users in their workflow (request for help, delegation, communication of personal findings, etc.) [33], [59]. Gestures and objects encode a lot of information in their shape, dynamics and direction, that can be directly interpreted in relation with the user, independently from the display output. Physical objects can also greatly improve coordination among actors for, e.g., handling priorities or assigning specific roles.

Relevant publications by team members this year: [9], [23], [16], [22] and major ones in recent years: [1], [10], [5], [3], [8].

4. Application Domains

4.1. Mission-critical systems

Mission-critical contexts of use include emergency response & management, and critical infrastructure operations, such as public transportation systems, communications and power distribution networks, or the operations of large scientific instruments such as particle accelerators and astronomical observatories. Central to these contexts of work is the notion of situation awareness [25], i.e., how workers perceive and understand elements of the environment with respect to time and space, such as maps and geolocated data feeds from the field, and how they form mental models that help them predict future states of those elements. One of the main challenges is how to best assist subject-matter experts in constructing correct mental models and making informed decisions, often under time pressure. This can be achieved by providing them with, or helping them efficiently identify and correlate, relevant and timely information extracted from large amounts of raw data, taking into account the often cooperative nature of their work and the need for task coordination. With this application area, our goal is to investigate novel ways of interacting with computing systems that improve collaborative data analysis capabilities and decision support assistance in a mission-critical, often time-constrained, work context.

Relevant publications by team members this year: [13], [19], [12].

4.2. Exploratory analysis of scientific data

Many scientific disciplines are increasingly data-driven, including astronomy, molecular biology, particle physics, or neuroanatomy. While making the right decision under time pressure is often less of critical issue when analyzing scientific data, at least not on the same temporal scale as truly time-critical systems, scientists are still faced with large-to-huge amounts of data. No matter their origin (experiments, remote observations, large-scale simulations), these data are difficult to understand and analyze in depth because of their sheer size and complexity. Challenges include how to help scientists freely-yet-efficiently explore their data, keep a trace of the multiple data processing paths they considered to verify their hypotheses and make it easy to backtrack, and how to relate observations made on different parts of the data and insights gained at different moments during the exploration process. With this application area, our goal is to investigate how data-centric interactive systems can improve collaborative scientific data exploration, where users' goals are more open-ended, and where roles, collaboration and coordination patterns [48] differ from those observed in mission-critical contexts of work.

Relevant publications by team members last year: [16], [24], [14], [18].

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Honorable mention for ActiveInk: (Th)Inking with Data at CHI 2019 .
- Best paper and best doctoral student paper awards for Influence of Color and Size of Particles on their Perceived Speed in Node-Link Diagrams at INTERACT 2019 .

BEST PAPERS AWARDS :

[22]
H. ROMAT, N. RICHE, K. HINCKLEY, B. LEE, C. APPERT, E. PIETRIGA, C. COLLINS.*ActiveInk: (Th)Inking with Data*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, CHI 2019 - Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019 [*DOI :* 10.1145/3290605.3300272], https://hal.archives-ouvertes.fr/hal-02020272

[20]
H. ROMAT, D. LEBOUT, E. PIETRIGA, C. APPERT.*Influence of Color and Size of Particles on Their Perceived Speed in Node-Link Diagrams*, in "INTERACT 2019 - 17th IFIP TC 13 International Conference on Human-Computer Interaction", Paphos, Cyprus, Springer, August 2019, p. 619-637 [*DOI :* 10.1007/978-3-030-29384-0_37], https://hal.inria.fr/hal-02274134

6. New Software and Platforms

6.1. Smarties

FUNCTIONAL DESCRIPTION: The Smarties system provides an easy way to add mobile interactive support to collaborative applications for wall displays.

It consists of (i) a mobile interface that runs on mobile devices for input, (ii) a communication protocol between the mobiles and the wall application, and (iii) libraries that implement the protocol and handle synchronization, locking and input conflicts. The library presents the input as an event loop with callback functions and handles all communication between mobiles and wall application. Developpers can customize the mobile interface from the wall application without modifying the mobile interface code.

On each mobile we find a set of cursor controllers associated with keyboards, widgets and clipboards. These controllers (pucks) can be shared by multiple collaborating users. They can control simple cursors on the wall application, or specific content (objects or groups of them). The developper can decide the types of widgets associated to pucks from the wall application side.

- Contact: Olivier Chapuis
- URL: http://smarties.lri.fr/

6.2. ZVTM

Zoomable Visual Transformation Machine

KEYWORDS: Big data - Visualization - Data visualization - Information visualization - Graph visualization

FUNCTIONAL DESCRIPTION: ZVTM is a toolkit enabling the implementation of multi-scale interfaces for interactively navigating in large datasets displayed as 2D graphics.

ZVTM is used for browsing large databases in multiple domains: geographical information systems, control rooms of complex facilitites, astronomy, power distribution systems.

The toolkit also enables the development of applications running on ultra-high-resolution wall-sized displays.

- Participants: Caroline Appert, Olivier Chapuis and Emmanuel Pietriga
- Contact: Emmanuel Pietriga
- Publications: Rapid Development of User Interfaces on Cluster-Driven Wall Displays with jBricks - A Toolkit for Addressing HCI Issues in Visual Language Environments
- URL: http://zvtm.sf.net

6.3. MapMosaic

KEYWORDS: Geo-visualization - Data visualization

SCIENTIFIC DESCRIPTION: GIS software applications and other mapping tools enable users to correlate data from multiple layers and gain insight from the resulting visualizations. However, most of these applications only feature basic, monolithic layer compositing techniques. These techniques do not always support users effectively in their tasks, as we observed during interviews with GIS experts. MapMosaic is a novel approach based on dynamic visual compositing that enables users to interactively create and manipulate local composites of multiple vector and raster map layers, taking into account the semantics and attribute values of objects and fields in the composit-ing process. We evaluated MapMosaic's interaction model against that of QGIS (a widely-used desktop GIS) and MAPublisher (a professional cartography tool) using the "Cognitive Dimensions" framework and through an analytical comparison, showing that MapMosaic's model is more flexible and can support users more effectively in their tasks.

FUNCTIONAL DESCRIPTION: MapMosaic is a novel approach to combine geographical layers based on dynamic visual compositing that enables users to interactively create and manipulate local composites of multiple vector and raster map layers. It takes into account the semantics and attribute values of objects and fields in the compositing process. MapMosaic aims at better supporting GIS users in their tasks such as correlating data from multiple layers and gaining insight from the resulting visualizations.

RELEASE FUNCTIONAL DESCRIPTION: First public release.

- Participants: Maria Jesus Lobo, Caroline Appert and Emmanuel Pietriga
- Contact: Emmanuel Pietriga
- Publications: MapMosaic: Dynamic Layer Compositing for Interactive Geovisualization - An Evaluation of Interactive Map Comparison Techniques
- URL: http://ilda.saclay.inria.fr/mapmuxing/mapmosaic/index.html

6.4. Baia

Before-and-after satellite image animation

KEYWORDS: Geo-visualization - 2D animation

SCIENTIFIC DESCRIPTION: Before-and-after image pairs show how entities in a given region have evolved over a specific period of time. Satellite images are a major source of such data, that capture how natural phenomena or human activity impact a geographical area. These images are used both for data analysis and to illustrate the resulting findings to diverse audiences. The simple techniques used to display them, including juxtaposing, swapping and monolithic blending, often fail to convey the underlying phenomenon in a meaningful manner. Baia is a framework to create advanced animated transitions, called animation plans, between before-and-after images. Baia relies on a pixel-based transition model that gives authors much expressive power, while keeping animations for common types of changes easy to create thanks to predefined animation primitives.

FUNCTIONAL DESCRIPTION: Baia is a framework to create advanced animated transitions, called animation plans, between before-and-after satellite images.

Before-and-after image pairs show how entities in a given region have evolved over a specific period of time. Satellite images are a major source of such data, that capture how natural phenomena or human activity impact a geographical area. These images are used both for data analysis and to illustrate the resulting findings to diverse audiences. The simple techniques used to display them, including juxtaposing, swapping and monolithic blending, often fail to convey the underlying phenomenon in a meaningful manner.

Baia relies on a pixel-based transition model that gives authors much expressive power. The animation editor enables authors to easily represent common types of changes thanks to predefined animation primitives and to sequence different changes across time.

RELEASE FUNCTIONAL DESCRIPTION: First public release

- Participants: Maria Jesus Lobo, Caroline Appert and Emmanuel Pietriga
- Contact: Emmanuel Pietriga
- Publication: Animation Plans for Before-and-After Satellite Images
- URL: http://ilda.saclay.inria.fr/mapmuxing/baia/index.html

6.5. LODAtlas

KEYWORDS: LOD - Linked open data - Semantic Web

SCIENTIFIC DESCRIPTION: The Web of Data is growing fast, as exemplified by the evolution of the Linked Open Data (LOD) cloud over the last ten years. One of the consequences of this growth is that it is becoming increasingly difficult for application developers and end-users to find the datasets that would be relevant to them. Semantic Web search engines, open data catalogs , datasets and frameworks such as LODStats and LOD Laundromat, are all useful but only give partial, even if complementary, views on what datasets are available on the Web. LODAtlas is a portal that enables users to find datasets of interest. Users can make different types of queries about both the datasets' metadata and contents, aggregated from multiple sources. They can then quickly evaluate the matching datasets' relevance, thanks to LODAtlas' summary visualizations of their general metadata, connections and contents.

FUNCTIONAL DESCRIPTION: The Web of Data is growing fast, as exemplified by the evolution of the Linked Open Data (LOD) cloud over the last ten years. One of the consequences of this growth is that it is becoming increasingly difficult for application developers and end-users to find the datasets that would be relevant to them. Semantic Web search engines, open data catalogs, datasets and frameworks such as LODStats and LOD Laundromat, are all useful but only give partial, even if complementary, views on what datasets are available on the Web. LODAtlas is a portal that enables users to find datasets of interest. Users can make different types of queries about both the datasets' metadata and contents, aggregated from multiple sources. They can then quickly evaluate the matching datasets' relevance, thanks to LODAtlas' summary visualizations of their general metadata, connections and contents.

- Participants: Caroline Appert, Marie Destandau, Ioana Manolescu, François Goasdoué, Sejla Cebiric, Hande Ozaygen and Emmanuel Pietriga
- Contact: Emmanuel Pietriga

- Publication: Browsing Linked Data Catalogs with LODAtlas
- URL: http://lodatlas.lri.fr

6.6. TouchTokens

KEYWORDS: Tangible interface - HCI

SCIENTIFIC DESCRIPTION: TouchTokens make it possible to easily build interfaces that combine tangible and gestural input using passive tokens and a regular multi-touch surface. The tokens constrain users' grasp, and thus, the relative spatial configuration of fingers on the surface, theoretically making it possible to design algorithms that can recognize the resulting touch patterns. See associated scientific articles below.

FUNCTIONAL DESCRIPTION: TouchTokens allow interface designers to build low-cost tangible interfaces. The technique consists in recognizing multi-touch patterns that are associated with specific passive tokens. Those physical tokens can be made out of any material to get tracked on any touch-sensitive surface. Implementations of the recognizer (in both TUIO and Android) and vector descriptions of the tokens ready for 3D-printing or laser-cutting are available

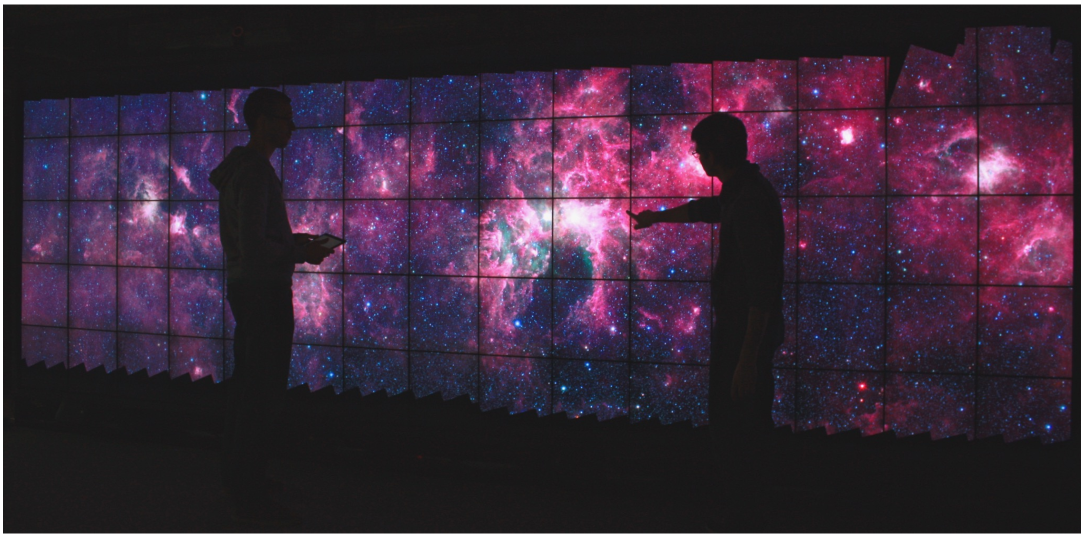
- Participants: Caroline Appert, Rafael Morales Gonzalez, Emmanuel Pietriga and Gilles Bailly
- Contact: Caroline Appert
- Publications: TouchTokens: Guiding Touch Patterns with Passive Tokens - Passive yet Expressive TouchTokens - Custom-made Tangible Interfaces with TouchTokens
- URL: https://www.lri.fr/~appert/touchtokens/

6.7. Platforms

6.7.1. Platform: WILDER

Ultra-high-resolution wall-sized displays [31] feature a very high pixel density over a large physical surface. Such platforms have properties that make them well-suited to the visualization of very large datasets. They can represent the data with a high level of detail while at the same time retaining context: users can transition from an overview of the data to a detailed view simply by physically moving in front of the wall display. Wall displays also offer good support for collaborative work, enabling multiple users to simultaneously visualize and interact with the displayed data. To make them interactive, wall-sized displays are increasingly coupled with input devices such as touch frames, motion-tracking systems and wireless multitouch devices, in order to enable multi-device and multi-user interaction with the displayed data. Application areas for such visualization platforms range from the monitoring of complex infrastructures and crisis management situations to tools for the exploratory visualization of scientific data.

WILDER is the latest ultra-high-resolution wall-sized display set up at Inria Saclay, and is one of the nodes of the Digiscope EquipEx. We use this platform for multiple projects, both fundamental HCI research, and research and development activities for specific application areas such as geographical informations systems and astronomy (Figure 2).



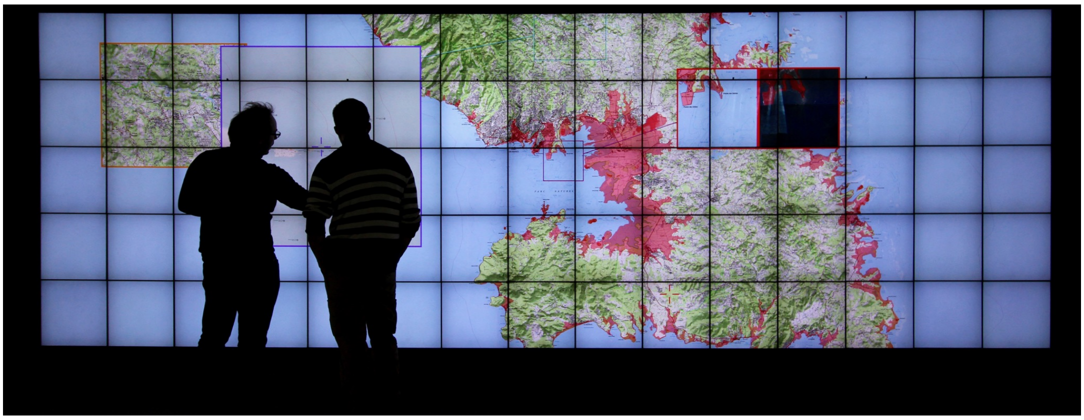


Figure 2. Example application area for ultra-high-resolution wall-sized displays: geographical information systems, and astronomical data analysis.

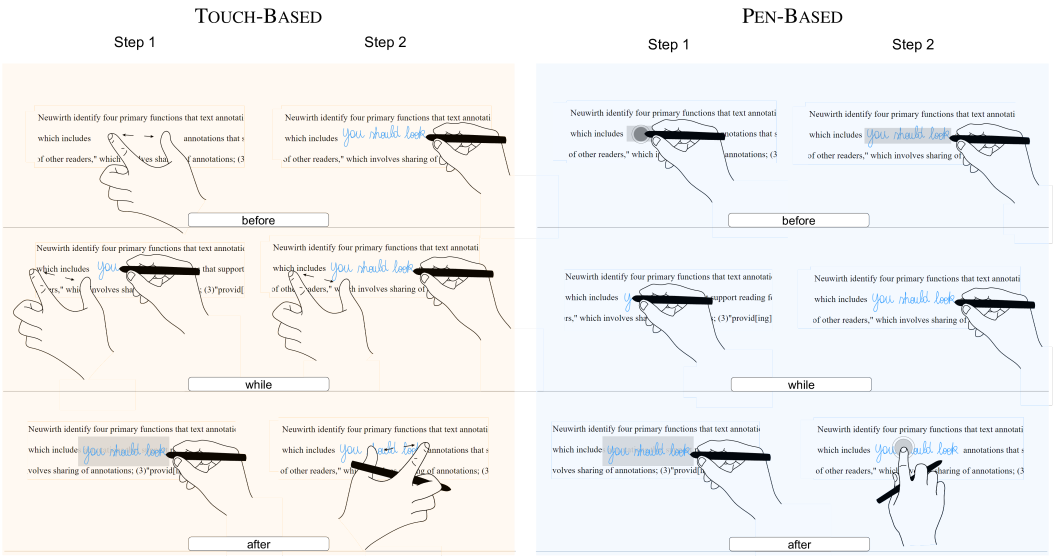


Figure 3. In the SpaceInk design space, both touch-based and pen-based techniques let users specify the strategy for creating white space at different moments: either before, while, or after annotating. With touch-based techniques, users are free to both gesture and write with a single hand (as illustrated in the after condition) or with two hands (as illustrated in the before condition).

7. New Results

7.1. Digital Ink and Data Manipulation

We investigated how pen and touch could be best combined to facilitate the digital annotation of documents. When editing or reviewing a document, people directly overlay ink marks on content. For instance, they underline words, or circle elements in a figure. These overlay marks often accompany in-context annotations in the form of handwritten footnotes and marginalia. People tend to put annotations close to the content that elicited them, but have to compose with the often-limited whitespace. Based on these observations, we explored a design space – which we call SpaceInk (UIST 2019 [9]) – of pen+touch techniques that make room for in-context annotations by dynamically reflowing documents. We identified representative techniques in this design space, spanning both new ones and existing ones, as illustrated in Figure 3. We evaluated them in a user study. The results of this study then informed the design of a prototype system which lets users concentrate on capturing fleeting thoughts, streamlining the overall annotation process by enabling the fluid inverleaving of space-making gestures with freeform ink.

Together with colleagues from the EPIC team at Microsoft Research (see Section 9.3.2.1), we also investigated the potential of digital inking for exploring heterogeneous datasets and trying to make sense of them. During sensemaking, people annotate insights: underlining sentences in a document or circling regions on a map. They jot down their hypotheses: drawing correlation lines on scatterplots or creating personal legends to track patterns. Based on these observations, we designed ActiveInk (CHI 2019 [22]), a system enabling people to seamlessly transition between exploring data and externalizing their thoughts using pen and touch as input channels. ActiveInk, illustrated in Figure 4, enables the natural use of pen for active reading behaviors, while supporting analytic actions by activating any of these ink strokes. Through a qualitative study with eight

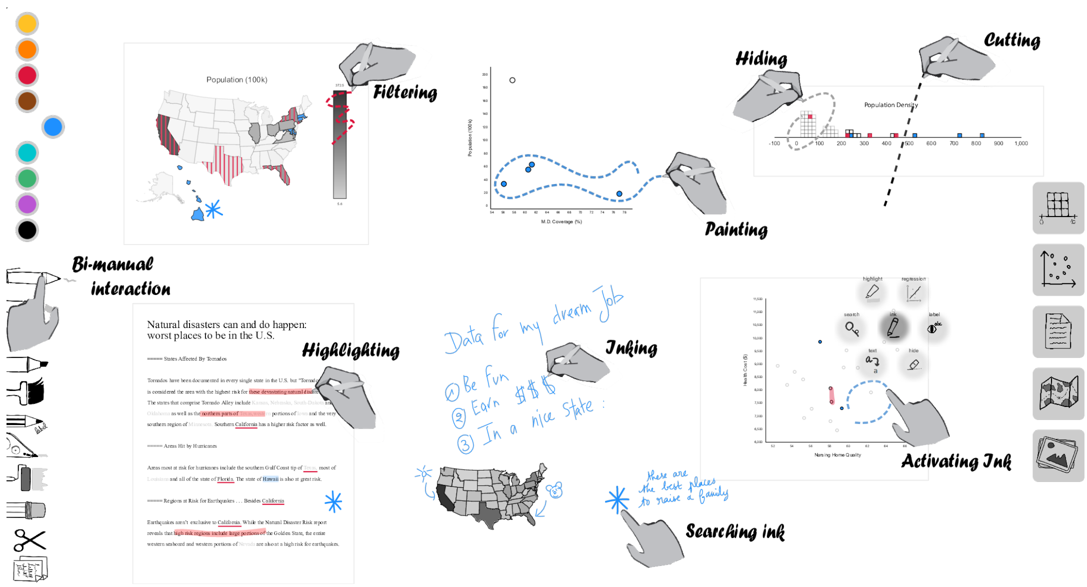


Figure 4. ActiveInk, a collaboration with Microsoft Research, affords smooth transition between using a digital pen for high-precision selections of heterogeneous data coming from multiple sources, and for externalizing thinking via notes and annotations. Ink strokes are leveraged to perform operations on underlying data.

participants, we observed active reading behaviors during data exploration and design principles to support sensemaking.

7.2. Novel Forms of Input in Immersive Environments

ILDA researchers have started to investigate input techniques for the specific context of immersive environments, based on, e.g., virtual or augmented reality. These hardware devices enable displaying large amounts of data in space to better support data analysis, and there is a growing group of research focusing on Immersive Analytics in the HCI community. The question of input for data manipulation in these environments is crucial, but challenging because users must be able to activate various commands or adjust various values while remaining free to move. In this context, using the whole body as an input device offers several advantages: 1) the body provides physical support as an interactive surface, which improves accuracy and makes it less tiring to interact; 2) using the body does not impair mobility and avoids handling devices; 3) proprioception makes it possible to interact eyes-free, including when choosing values in a range; 4) by leveraging spatial memory, the body helps memorizing commands, thus interacting in expert mode (i.e., perform quick actions without visual feedback). ILDA team members participated to a position paper on this topic, which analyzes various ways of interacting with the body, discussing their pros and cons as well as associated challenges for immersive analytics [23].

7.3. Multivariate Network Visualization

Edges in networks often represent transfer relationships between vertices. When visualizing such networks as node-link diagrams, animated particles flowing along the links can effectively convey this notion of transfer. Variables that govern the motion of particles, their speed in particular, may be used to visually represent edge data attributes. Few guidelines exist to inform the design of these particle-based network visualizations, however. Following up on our initial investigation of motion as an encoding channel for edge attributes in

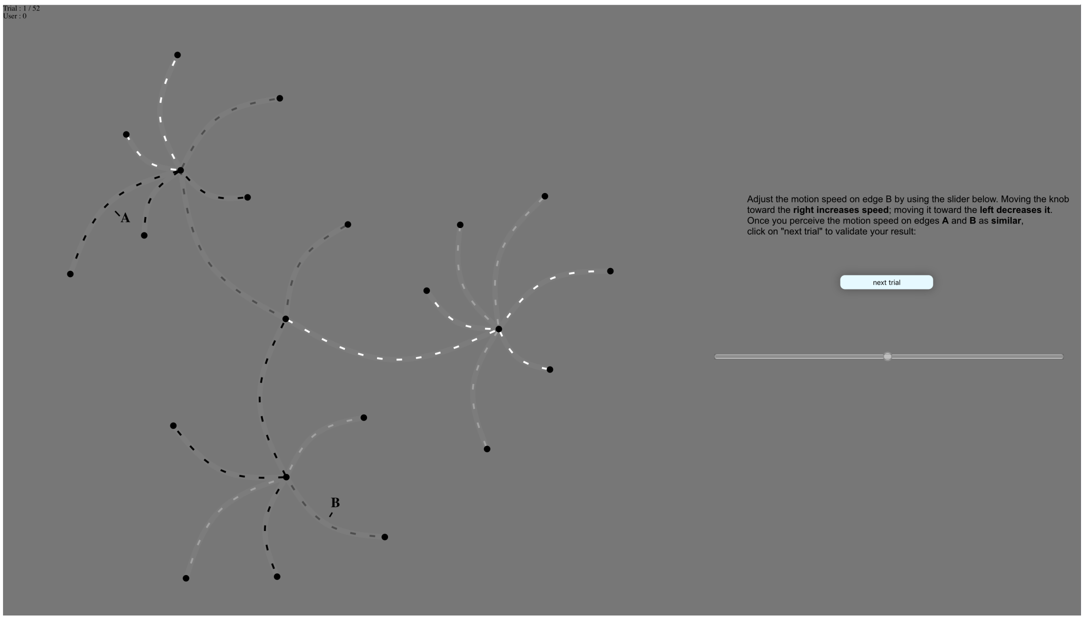


Figure 5. Experimental task used in the study about the influence of color and size of moving particles on their perceived speed in node-link diagrams. Participants had to adjust the speed of particles on a specific edge so that it would match that of particles on another edge, using a slider.

multivariate network visualization [75], we investigated the influence of color and size of moving particles on their perceived speed in node-link diagrams (INTERACT 2019 [20]). Empirical studies so far had only looked at the different motion variables in isolation, independently from other visual variables controlling the appearance of particles, such as their color or size. We ran a study of the influence of several visual variables on users' perception of the speed of particles. Part of the experimental setup is illustrated in Figure 5. Our results show that particles' luminance, chromaticity and width do not interfere with their perceived speed. But variations in their length make it more difficult for users to compare the relative speed of particles across edges.

Figure 6. Graphies running on a tablet with support for pen+touch. Graphies is developed entirely using Web technologies.

Beyond questions of perception of information in multivariate network visualizations, we also investigated the problem of creating environments for the design of multivariate network visualizations, with a focus on expressive design. Expressive design environments enable visualization designers not only to specify chart types and visual mappings, but also to customize individual graphical marks, as they would in a vector graphics drawing tool. Prior work had mainly investigated how to support the expressive design of a wide range of charts generated from tabular data: bar charts, scatterplots, maps, *etc*. But multivariate network data structures raise specific challenges and opportunities in terms of visual design and interactive authoring. Together with company TKM (see Section 8.1), we developed an expressive design environment for node-link diagrams generated from multivariate networks called Graphies (TVCG 2019, [15]), illustrated in Figure 6. We followed a user-centered design approach, involving expert analysts from TKM, and validated the approach through a study in which participants successfully reproduced several expressive designs, and created their own designs as well.

7.4. Visualization in Specific Application Areas

Finally, we worked in collaboration with other researchers on projects aimed at investigating how visualization can support experts in different application areas.

In the area of geovisualization, we performed a comparison of visualization techniques to help analysts identify correlation between variables over space and time (TVCG/InfoVis 2019, [6]). Observing the relationship between two or more variables over space and time is essential in many application domains. For instance, looking, for different countries, at the evolution of both the life expectancy at birth and the fertility rate will give an overview of their demographics. The choice of visual representation for such multivariate data is key to enabling analysts to extract patterns and trends. We conducted a study comparing three techniques that are representative of different strategies to visualize geo-temporal multivariate data. Participants performed a series of tasks that required them to identify if two variables were correlated over time and if there was a pattern in their evolution. Our results showed that a visualization's effectiveness depends strongly on the task to be carried out. Based on this study's findings, we derived a set of design guidelines about geo-temporal visualization techniques for communicating correlation.

Together with researchers from INRA, we performed an exploratory study about the visual exploration of model simulations for a range of experts (CHI 2019, [16]). Experts in different domains rely increasingly on simulation models of complex processes to reach insights, make decisions, and plan future projects. These models are often used to study possible trade-offs, as experts try to optimize multiple conflicting objectives in a single investigation. Understanding all the model intricacies, however, is challenging for a single domain expert. This project introduced a simple approach to support multiple experts when exploring complex model results, working concurrently on a shared visualization surface. The results of an observational study focusing on the link between expertise and insight generation during the analysis process, revealed the different exploration strategies and multi-storyline approaches that domain experts adopt during trade-off analysis. This eventually led to recommendations for collaborative model exploration systems.

We collaborated with researchers in databases from Université Paris Descartes on progressive similarity search on time-series data (BigVis 2019 workshop, [24]). Time-series data are increasing at a dramatic rate, yet their analysis remains highly relevant in a wide range of human activities. Due to their volume, existing systems dealing with time-series data cannot guarantee interactive response times, even for fundamental tasks such as similarity search. This paper presented our vision to develop analytic approaches that support exploration and decision making by providing progressive results, before the final and exact ones have been computed. Findings from our experiment indicated that there is a gap between the time the most similar answer is found and the time when the search algorithm terminates, resulting in inflated waiting times without any improvement. These findings led to preliminary ideas about computing probabilistic estimates of the final results that could help users decide when to stop the search process.

In the field of Education, we contributed to EduClust, an online visualization application for teaching clustering algorithms (EuroGraphics 2019, [18]). EduClust combines visualizations, interactions, and animations to facilitate the understanding and teaching of clustering steps, parameters, and procedures. Traditional classroom settings aim for cognitive processes like remembering and understanding. We designed EduClust for expanded educational objectives like applying and evaluating. The application can be used by both educators to prepare teaching material and examples, and by students to explore clustering differences and discover algorithmic subtleties.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Tecknowmetrix (TKM): ANRT/CIFRE PhD (Hugo Romat), 3 years, June 2016-August 2019.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Inria Project Lab (IPL)

ILDA participates to Inria Project Lab iCODA : Data Journalism : knowledge-mediated Content and Data Interactive Analytics, that started in 2017. A key issue in data science is the design of algorithms that enable analysts to infer information and knowledge by exploring heterogeneous information sources, structured data, or unstructured content. With journalism data as a landmark use-case, iCODA aims to develop the scientific and technological foundation for collaborative, heterogeneous data analysis, guided by formalized, user-centric knowledge. The project relies on realistic scenarios in data-journalism to assess the contribution of the project to this area. iCODA is at the crossroads of several research areas (content analysis, data management, knowledge representation, visualization) and is part of a club of partners of the world of the press. Equipes-projets Inria : Graphik, Ilda, Linkmedia, Cedar. Press partners: Le Monde, OuestFrance, AFP. Participants: Anastasia Bezerianos (PI), Emmanuel Pietriga, Tong Xue, Vanessa Peña-Araya, Nicole Barbosa Sultanum.

9.2. European Initiatives

9.2.1. Collaborations with Major European Organizations

Deutsches Elektronen-Synchrotron (DESY): Scientific collaboration on the design and implementation of user interfaces for array operations monitoring and control for the Cherenkov Telescope Array (CTA) project, to be built in the Canary Islands (Spain) and in the Atacama desert (Chile), 2 years, contract started May 2018.

9.3. International Initiatives

9.3.1. Inria International Labs

Inria Chile. From 2012 to 2015, Emmanuel Pietriga was the scientific leader of the Massive Data team at Inria Chile, working on projects in collaboration with the ALMA radio-telescope and the Millenium Institute of Astrophysics. He is now scientific advisor to Inria Chile's visualization projects, and is actively involved in the collaboration between Inria Chile and the LSST on the design and development of user interfaces for operations monitoring and control.

9.3.2. Inria International Partners

Association of Universities for Research in Astronomy (AURA): contract, jointly with Inria Chile, on the design and implementation of user interfaces for telescope operations monitoring and control for the Large Synoptic Survey Telescope (LSST) project, under construction in the Atacama desert (Chile), started 2017. Participants: Emmanuel Pietriga (ILDA), Sebastian Fehlandt (Inria Chile), José Galaz (Inria Chile), Sebastian Pereira (Inria Chile), Grazia Prato (Inria Chile).

9.3.2.1. Informal International Partners

We have had multiple collaboration projects with Microsoft Research in Redmond, USA. Hugo Romat visited the EPIC team for three months, and this collaboration led to the following publications at CHI 2018 [75], CHI 2019 [22] and UIST 2019 [9]. Anastasia Bezerianos also continues working with that team on topics related to smartwatch interaction and visualization that appeared in TVCG 2019 (InfoVis 2018) [11].

Our long-term collaboration with University of Konstanz, Germany continues. After publications at TVCG/InfoVis in 2014 and 2018 [46], [47], Anastasia Bezerianos has co-authored a paper at Eurographics 2019 with these colleagues [18].

Finally, our ongoing collaboration with Northwestern University, USA continues. Anastasia Bezerianos and past PhD student Evanthia Dimara (PhD defended in 2017) have worked on publications in TVCG 2019 [12] [42].

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- ACM CHI Steering Committee, ACM SIGCHI Conference on Human Factors in Computing Systems: Caroline Appert

10.1.1.2. Member of the Organizing Committees

- IEEE VIS (IEEE Visualization Conference) 2019, 2020: Anastasia Bezerianos (Organization Committee - Workshops co-Chair)

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- CHI 2020, 38th ACM SIGCHI Conference on Human Factors in Computing Systems: Anastasia Bezerianos (SC - subcommittee chair)

10.1.2.2. Member of the Conference Program Committees

- CHI 2020, 38th ACM SIGCHI Conference on Human Factors in Computing Systems: Caroline Appert, Emmanuel Pietriga
- IEEE VIS 2019, Visualization Conference (InfoVis): Anastasia Bezerianos
- TheWebConf (WWW) 2020, 29th Web Conference, research track Web Mining and Content Analysis: Emmanuel Pietriga
- ESWC 2020, 17th Extended Semantic Web Conference: Emmanuel Pietriga
- TheWebConf (WWW) 2019, 28th Web Conference, research track Web Content Analysis, Semantics, and Knowledge: Emmanuel Pietriga
- ESWC 2019, 16th Extended Semantic Web Conference: Emmanuel Pietriga
- EICS 2019, 11th ACM Symposium on Engineering Interactive Computing Systems, Tech Notes, Olivier Chapuis

10.1.2.3. Reviewer

- ACM EICS 2020, SIGCHI Symposium on Engineering Interactive Computing Systems: Emmanuel Pietriga
- ACM CHI 2019, Conference on Human Factors in Computing Systems: Caroline Appert, Olivier Chapuis, Emmanuel Pietriga, Hugo Romat.
- ACM CHI 2020, Conference on Human Factors in Computing Systems: Olivier Chapuis, Vanessa Peña-Araya
- ACM UIST 2019, Interface Software and Technologies Symposium: Olivier Chapuis, Emmanuel Pietriga, Hugo Romat.
- ACM DIS 2019, Designing Interactive Systems: Caroline Appert
- ACM MobileHCI 2019, International Conference on Human-Computer Interaction with Mobile Devices and Services: Caroline Appert, Olivier Chapuis
- IEEE VIS 2019, Visualization Conference (InfoVis): Emmanuel Pietriga, Vanessa Peña-Araya (short papers)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- ACM ToCHI, Transactions on Computer-Human Interaction: Caroline Appert (associate editor)

10.1.3.2. Reviewer - Reviewing Activities

- ACM ToCHI, Transactions on Computer-Human Interaction: Olivier Chapuis
- IEEE TVCG, Transactions on Visualization and Computer graphics: Anastasia Bezerianos

10.1.4. Invited Talks

- Caroline Appert: Lowering the cost of transitions between devices, Microsoft Research Faculty Summit, Seattle, USA, July 2019.
- Anastasia Bezerianos: Interactive visualizations and large displays, Indiana University, USA, December 2019.
- Olivier Chapuis: Multi-user interaction on wall displays, Public lectures on interactive design, Chalmers University, Göteborg, Sweden, November 2019.
- Emmanuel Pietriga: Créativité et visualisation de réseaux sous formes de diagrammes noeuds-liens, École d'été de cartographie et visualisation, Lyon, France, July 2019.

10.1.5. Scientific Expertise

- Evaluator for NSERC Discovery Grants (Canada): Anastasia Bezerianos, Olivier Chapuis

10.1.6. Research Administration

- Deputy Director of the Laboratoire de Recherche en Informatique (LRI): Olivier Chapuis
- President of Inria Saclay - Île de France's Commission for Technological Development (CDT): Emmanuel Pietriga
- Deputy head of Pôle "Données, Connaissances, Apprentissage et Interaction" at École Doctorale Paris Saclay (ED STIC): Caroline Appert
- Member of Conseil de Département (LRI): Anastasia Bezerianos
- Member of Commissions Consultatives de Spécialités d'Université (CCSU) at Université Paris-Sud/Paris Saclay: Anastasia Bezerianos

10.1.7. Learned societies

- Association Francophone d'Interaction Homme-Machine (AFIHM), in charge of the relation with the SIF: Olivier Chapuis.
- SigCHI Paris Local Chapter, chair: Caroline Appert.

10.1.8. Hiring committees

- MCF Hiring committee, Univ. Paris Sud - LIMSI: Caroline Appert
- MCF Hiring committee, Univ. Technologique Troyes - Tech-CICO: Caroline Appert

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Ingénieur: Emmanuel Pietriga, Data Visualization (INF552), 36h, 3A/M1, École Polytechnique.

Ingénieur: Caroline Appert, Data Visualization (INF552), 18h, 3A/M1, École Polytechnique.

Ingénieur: Vanessa Peña-Araya, Data Visualization (INF552), 16h, 3A/M1, École Polytechnique.

Master: Emmanuel Pietriga, Data Visualization, 24h, M2 Informatique Décisionnelle, Univ. Paris-Dauphine.

Master: Caroline Appert, Evaluation of Interactive Systems (Advanced), 21h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Evaluation of Interactive Systems (Intro), 16h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Programming of Interactive Systems, 21h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Career Seminar, 21h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Augmented Reality and Tangible Interaction, 12h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Interactive Information Visualization, 10.5h, M2 Interaction, Univ. Paris-Sud.

Master: Anastasia Bezerianos, HCI Project, 21h, M2 Interaction and HCID, Univ. Paris-Sud.

Master: Anastasia Bezerianos, Intro aux Systèmes Interactives, 15h, M1 Informatique, Univ. Paris-Sud.

Master: Eugénie Brasier, Programming of Interactive Systems, 10h30, M2 Interaction and HCID, Univ. Paris-Saclay.

Master: Eugénie Brasier, Career Seminar, 10h30, M2 Interaction and HCID, Univ. Paris-Saclay.

Master: Emmanuel Courtoux, Fundamentals of HCI, 21h, M2 Interaction and HCID, Univ. Paris-Saclay.

Master: Emmanuel Courtoux, Digital Fabrication, 21h, M2 Interaction and HCID, Univ. Paris-Saclay.

Master: Raphaël James, Programming of Interactive Systems, 12h, Univ. Paris-Saclay.

Master: Raphaël James, Fundamentals of HCI, 21h, Univ. Paris-Saclay.

Master: Tong Xue, Design of Interactive Systems, 21h, Univ. Paris-Saclay.

Licence: Eugénie Brasier, Algorithmique, 18h, PeiP2, Polytech.

License: Raphaël James, Algorithm/C/C++, 2h, Polytech.

10.2.2. Supervision

PhD in progress : Emmanuel Courtoux, Tangible Collaborative Interaction for Wall-sized Displays, since October 2019, Advisors: Olivier Chapuis, Caroline Appert

PhD in progress : Eugénie Brasier, Interaction techniques for remote manipulation in multi-display environments, since October 2018, Advisors: Caroline Appert

PhD in progress : Raphaël James, Environnements de réalité physique et augmentée utilisés dans l'analyse visuelle collaborative, since October 2018, Advisors: Anastasia Bezerianos, Olivier Chapuis, Tim Dwyer

PhD in progress : Tong Xue, Interactive Visualization for Data Journalism, since October 2018, Advisors: Anastasia Bezerianos, Emmanuel Pietriga

PhD in progress : Marie Destandau, Interactive Visual Exploration of Webs of Data, since October 2017, Advisors: Emmanuel Pietriga

PhD (defended) : Anna Gogolou, Iterative and expressive querying for big data series, October 2016 - November 2019, Advisors: Anastasia Bezerianos, Themis Palpanas

PhD (defended [76]) : Hugo Romat, Visual exploration and interactive manipulation techniques for collections of heterogeneous data and documents, June 2016 - November 2019, Advisors: Caroline Appert, Emmanuel Pietriga

PhD Internship: Nicole Barbosa Sultanum (from University of Toronto), Navigation and exploration of multivariate and dynamic graphs, 3 months, Advisor: Anastasia Bezerianos

Master 2 Internship: Marzieh Rafiei, Wall displays in Virtual Reality, 6 months, Advisors: Olivier Chapuis

Master 2 Internship: Emmanuel Courtoux, Tangible interaction for wall displays, 6 months, Advisors: Olivier Chapuis, Caroline Appert

10.2.3. Juries

PhD: Thibault Raffaillac, Université de Lille: Emmanuel Pietriga (examinateur)

PhD: Yujiro Okuya, Université Paris-Sud: Caroline Appert (president)

PhD: Gaëlle Richer, Université de Toulouse: Caroline Appert (reviewer)

PhD: Khanh Duy Le, Chalmers University at Göteborg: Olivier Chapuis (reviewer)

10.3. Popularization

10.3.1. Articles and contents

- 2019-06-29: France 2 (TV) - 20h30 le samedi, Le premier pas sur la lune, https://www.france.tv/france-2/20h30-le-samedi/1016245-le-premier-pas-sur-la-lune.html
- 2019-06-12: Ciel & Espace - Sanctuary, le projet français qui vise à faire de la Lune un lieu de la mémoire humaine, https://www.cieletespace.fr/actualites/sanctuary-le-projet-francais-qui-vise-a-faire-de-la-lune-un-lieu-de-la-memoire-humaine
- 2019-05-13: France 2 (TV) - Télématin, Sanctuary (11'30-15'30), https://www.france.tv/france-2/telematin/981641-telematin.html
- 2019-04-22: France 3 (TV) - Soir 3, Humanité : la Lune comme sanctuaire, https://www.francetvinfo.fr/sciences/espace/humanite-la-lune-comme-sanctuaire_3411135.html
- 2019-03-29: Science & Avenir - Sanctuary : le projet fou d'envoyer un selfie ou sa dédicace sur la Lune, https://www.sciencesetavenir.fr/espace/systeme-solaire/sanctuary-le-projet-fou-d-envoyer-un-temoignage-artistique-et-scientifique-sur-la-lune_122541
- 2019-03-18: Inria.fr - Sanctuary : une trace sur la Lune pour le cinquantenaire de la mission Apollo 11, https://www.inria.fr/fr/sanctuary-pour-laisser-une-trace-sur-la-lune
- 2019-03-16: Ouest France - Climat. Ce que les Français font déjà, ce qu'ils se déclarent prêts à faire (Storifier), https://www.ouest-france.fr/politique/grand-debat-national/climat-ce-que-les-francais-font-deja-ce-qu-ils-se-declarent-prets-faire-6264587
- 2019-02-15: Métro parisien, La fresque du couloir de la station Montparnasse-Bienvenüe (ZRaildar), http://80ans.cnrs.fr/ressource/la-fresque/

11. Bibliography

Major publications by the team in recent years

[1] B. FRUCHARD, E. LECOLINET, O. CHAPUIS. *MarkPad: Augmenting Touchpads for Command Selection*, in "Proceedings of the 35th international conference on Human factors in computing systems", Denver, United States, CHI '17, May 2017 [*DOI :* 10.1145/3025453.3025486], https://hal.archives-ouvertes.fr/hal-01437093

[2] A. GOGOLOU, T. TSANDILAS, T. PALPANAS, A. BEZERIANOS. *Comparing Similarity Perception in Time Series Visualizations*, in "IEEE Transactions on Visualization and Computer Graphics", October 2018, 11, https://hal.inria.fr/hal-01845008

[3] C. LIU, O. CHAPUIS, M. BEAUDOUIN-LAFON, E. LECOLINET. *CoReach: Cooperative Gestures for Data Manipulation on Wall-sized Displays*, in "Proceedings of the 35th international conference on Human factors in computing systems", Denver, United States, CHI '17, ACM, May 2017 [*DOI :* 10.1145/3025453.3025594], https://hal.archives-ouvertes.fr/hal-01437091

[4] M.-J. LOBO, C. APPERT, E. PIETRIGA.*Animation Plans for Before-and-After Satellite Images*, in "IEEE Transactions on Visualization and Computer Graphics", January 2018, vol. 24, forthcoming [*DOI :* 10.1109/TVCG.2018.2796557], https://hal.inria.fr/hal-01773882

[5] R. MORALES GONZÁLEZ, C. APPERT, G. BAILLY, E. PIETRIGA.*TouchTokens: Guiding Touch Patterns with Passive Tokens*, in "2016 CHI Conference on Human Factors in Computing Systems", San Jose, CA, United States, May 2016 [*DOI :* 10.1145/2858036.2858041], https://hal.archives-ouvertes.fr/hal-01315130

[6] V. PEÑA-ARAYA, E. PIETRIGA, A. BEZERIANOS.*A Comparison of Visualizations for Identifying Correlation over Space and Time*, in "IEEE Transactions on Visualization and Computer Graphics", October 2019 [*DOI :* 10.1109/TVCG.2019.2934807], https://hal.archives-ouvertes.fr/hal-02320617

[7] E. PIETRIGA, H. GÖZÜKAN, C. APPERT, M. DESTANDAU, Š. ČEBIRIĆ, F. GOASDOUÉ, I. MANOLESCU.*Browsing Linked Data Catalogs with LODAtlas*, in "ISWC 2018 - 17th International Semantic Web Conference", Monterey, United States, Springer, October 2018, p. 137-153, https://hal.inria.fr/hal-01827766

[8] A. PROUZEAU, A. BEZERIANOS, O. CHAPUIS.*Evaluating Multi-User Selection for Exploring Graph Topology on Wall-Displays*, in "IEEE Transactions on Visualization and Computer Graphics", August 2017, vol. 23, n^o^ 8, p. 1936–1951 [*DOI :* 10.1109/TVCG.2016.2592906], https://hal.archives-ouvertes.fr/hal-01348578

[9] H. ROMAT, E. PIETRIGA, N. HENRY-RICHE, K. HINCKLEY, C. APPERT.*SpaceInk: Making Space for In-Context Annotations*, in "UIST 2019 - 32nd ACM User Interface Software and Technology", Nouvelle-Orleans, United States, October 2019 [*DOI :* 10.1145/3332165.3347934], https://hal.inria.fr/hal-02286161

[10] V. RUSŇÁK, C. APPERT, O. CHAPUIS, E. PIETRIGA.*Designing Coherent Gesture Sets for Multi-scale Navigation on Tabletops*, in "Proceedings of the 36th international conference on Human factors in computing systems", Montreal, Canada, CHI '18, ACM, April 2018, p. 142:1-142:12 [*DOI :* 10.1145/3173574.3173716], https://hal.inria.fr/hal-01722189

Publications of the year

Articles in International Peer-Reviewed Journal

[11] T. BLASCHECK, L. BESANÇON, A. BEZERIANOS, B. LEE, P. ISENBERG.*Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches*, in "IEEE Transactions on Visualization and Computer Graphics", January 2019, vol. 25, n^o^ 1, p. 616–629 [*DOI :* 10.1109/TVCG.2018.2865142], https://hal.inria.fr/hal-01851306

[12] E. DIMARA, S. FRANCONERI, C. PLAISANT, A. BEZERIANOS, P. DRAGICEVIC.*A Task-based Taxonomy of Cognitive Biases for Information Visualization*, in "IEEE Transactions on Visualization and Computer Graphics", 2019, forthcoming [*DOI :* 10.1109/TVCG.2018.2872577], https://hal.sorbonne-universite.fr/hal-01868738

[13] M.-J. LOBO, C. APPERT, E. PIETRIGA.*Animation Plans for Before-and-After Satellite Images*, in "IEEE Transactions on Visualization and Computer Graphics", February 2019, vol. 25, n^o^ 2, p. 1347-1360 [*DOI :* 10.1109/TVCG.2018.2796557], https://hal.inria.fr/hal-01773882

[14] V. Peña-Araya, E. Pietriga, A. Bezerianos.*A Comparison of Visualizations for Identifying Correlation over Space and Time*, in "IEEE Transactions on Visualization and Computer Graphics", October 2019 [*DOI :* 10.1109/TVCG.2019.2934807], https://hal.archives-ouvertes.fr/hal-02320617

[15] H. Romat, C. Appert, E. Pietriga.*Expressive Authoring of Node-Link Diagrams with Graphies*, in "IEEE Transactions on Visualization and Computer Graphics", November 2019, 1 [*DOI :* 10.1109/TVCG.2019.2950932], https://hal.inria.fr/hal-02358999

International Conferences with Proceedings

[16] N. Boukhelifa, A. Bezerianos, I. Cristian Trelea, N. Perrot, E. Lutton.*An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, May 2019 [*DOI :* 10.1145/3290605], https://hal.inria.fr/hal-02005699

[17] M. Destandau.*Interactive Visualisation Techniques for the Web of Data*, in "Companion The 2019 World Wide Web Conference", San Francisco, United States, ACM Press, 2019, p. 17-21 [*DOI :* 10.1145/3308560.3314189], https://hal.archives-ouvertes.fr/hal-02416185

[18] J. Fuchs, P. Isenberg, A. Bezerianos, M. Miller, D. Keim.*EduClust -A Visualization Application for Teaching Clustering Algorithms*, in "Eurographics - 40th Annual Conference of the European Association for Computer Graphics", Genova, Italy, May 2019, p. 1-8, Education paper, https://hal.inria.fr/hal-02103293

[19] I. Oya, E. Antolini, M. Füssling, J. A. Hinton, A. Mitchell, S. Schlenstedt, L. Baroncelli, A. Bulgarelli, V. Conforti, N. Parmiggiani, J. Borkowski, A. Carosi, J. Jacquemier, G. Maurin, J. Colomé, C. Hoischen, D. Lebout, E. Lyard, R. Walter, D. Melkumyan, K. Mosshammer, I. Sadeh, T. Schmidt, P. Wegner, U. Schwanke, J. Schwarz, G. Tosti.*The Array Control and Data Acquisition System of the Cherenkov Telescope Array*, in "Proceedings of ICALEPCS 2019", New York, NY, United States, October 2019, https://hal.inria.fr/hal-02421340

[20] *Best Paper*
H. Romat, D. Lebout, E. Pietriga, C. Appert.*Influence of Color and Size of Particles on Their Perceived Speed in Node-Link Diagrams*, in "INTERACT 2019 - 17th IFIP TC 13 International Conference on Human-Computer Interaction", Paphos, Cyprus, Springer, August 2019, p. 619-637 [*DOI :* 10.1007/978-3-030-29384-0_37], https://hal.inria.fr/hal-02274134.

[21] H. Romat, E. Pietriga, N. Henry-Riche, K. Hinckley, C. Appert.*SpaceInk: Making Space for In-Context Annotations*, in "UIST 2019 - 32nd ACM User Interface Software and Technology", Nouvelle-Orleans, United States, October 2019 [*DOI :* 10.1145/3332165.3347934], https://hal.inria.fr/hal-02286161

[22] *Best Paper*
H. Romat, N. Riche, K. Hinckley, B. Lee, C. Appert, E. Pietriga, C. Collins.*ActiveInk: (Th)Inking with Data*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems", Glasgow, United Kingdom, CHI 2019 - Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019 [*DOI :* 10.1145/3290605.3300272], https://hal.archives-ouvertes.fr/hal-02020272.

Conferences without Proceedings

[23] B. FRUCHARD, A. PROUZEAU, O. CHAPUIS, E. LECOLINET.*Leveraging Body Interactions to Support Immersive Analytics*, in "CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems - Workshop on Interaction Design & Prototyping for Immersive Analytics", Glasgow, United Kingdom, May 2019, 10 pages, https://hal.archives-ouvertes.fr/hal-02095993

[24] A. GOGOLOU, T. TSANDILAS, T. PALPANAS, A. BEZERIANOS.*Progressive Similarity Search on Time Series Data*, in "BigVis 2019 - 2nd International Workshop on Big Data Visual Exploration and Analytics", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-02103998

References in notes

[25] M. R. ENDSLEY, D. G. JONES (editors). *Designing for Situation Awareness: an Approach to User-Centered Design*, CRC Press, Taylor & Francis, 2012

[26] S. ABITEBOUL, P. BUNEMAN, D. SUCIU.*Data on the Web: From Relations to Semistructured Data and XML*, Morgan Kaufmann, 1999

[27] S. ABITEBOUL, P. SENELLART, V. VIANU.*The ERC Webdam on Foundations of Web Data Management*, in "Proc. of the 21st International Conference Companion on World Wide Web", WWW '12 Companion, ACM, 2012, p. 211–214, http://doi.acm.org/10.1145/2187980.2188011

[28] J. ALVINA, C. APPERT, O. CHAPUIS, E. PIETRIGA.*RouteLens: Easy Route Following for Map Applications*, in "Proceedings of the International Working Conference on Advanced Visual Interfaces", AVI '14, ACM, 2014, p. 125–128 [*DOI :* 10.1145/2598153.2598200]

[29] C. APPERT, O. CHAPUIS, E. PIETRIGA.*High-precision magnification lenses*, in "Proceedings of the 28th international conference on Human factors in computing systems", CHI '10, ACM, 2010, p. 273–282 [*DOI :* 10.1145/1753326.1753366]

[30] C. APPERT, J.-D. FEKETE.*OrthoZoom Scroller: 1D Multi-scale Navigation*, in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems", New York, NY, USA, CHI '06, ACM, 2006, p. 21–30, http://doi.acm.org/10.1145/1124772.1124776

[31] M. BEAUDOUIN-LAFON, S. HUOT, M. NANCEL, W. MACKAY, E. PIETRIGA, R. PRIMET, J. WAGNER, O. CHAPUIS, C. PILLIAS, J. R. EAGAN, T. GJERLUFSEN, C. KLOKMOSE.*Multisurface Interaction in the WILD Room*, in "IEEE Computer", 2012, vol. 45, n° 4, p. 48-56 [*DOI :* 10.1109/MC.2012.110]

[32] E. BENSON, D. KARGER.*End-users Publishing Structured Information on the Web: An Observational Study of What, Why, and How*, in "Proc. of the SIGCHI Conference on Human Factors in Computing Systems", CHI '14, ACM, 2014, p. 1265–1274, http://doi.acm.org/10.1145/2556288.2557036

[33] J. BERNDTSSON, M. NORMARK.*The coordinative functions of flight strips: air traffic control work revisited*, in "Proceedings of the international ACM SIGGROUP conference on Supporting group work", GROUP '99, ACM, 1999, p. 101–110, http://doi.acm.org/10.1145/320297.320308

[34] T. BERNERS-LEE, Y. CHEN, L. CHILTON, D. CONNOLLY, R. DHANARAJ, J. HOLLENBACH, A. LERER, D. SHEETS.*Tabulator: Exploring and Analyzing linked data on the Semantic Web*, in "Proc. of the International Semantic Web User Interaction Workshop", 2006

[35] J. BERTIN.*Semiology of Graphics*, University of Wisconsin Press, 1983

[36] A. BEZERIANOS, R. BALAKRISHNAN.*View and Space Management on Large Displays*, in "IEEE Computer Graphics and Applications", 2005, vol. 25, n^o 4, p. 34-43

[37] A. BEZERIANOS.*Using alternative views for layout, comparison and context switching tasks in wall displays*, in "OZCHI", 2007, p. 303-310

[38] D. BONNET, C. APPERT, M. BEAUDOUIN-LAFON.*Extending the Vocabulary of Touch Events with ThumbRock*, in "Proceedings of Graphics Interface", GI '13, CIPS, 2013, p. 221-228, http://dl.acm.org/citation.cfm?id=2532129.2532166

[39] O. CHAPUIS, A. BEZERIANOS, S. FRANTZESKAKIS.*Smarties: An Input System for Wall Display Development*, in "CHI '14", Toronto, Canada, ACM, April 2014, p. 2763-2772 [*DOI :* 10.1145/2556288.2556956], https://hal.archives-ouvertes.fr/hal-00979034

[40] A. COLLINS, A. BEZERIANOS, G. MCEWAN, M. RITTENBRUCH, R. WASINGER, J. KAY.*Understanding File Access Mechanisms for Embedded Ubicomp Collaboration Interfaces*, in "Proceedings of the 11th International Conference on Ubiquitous Computing", New York, NY, USA, UbiComp '09, ACM, 2009, p. 135–144, http://doi.acm.org/10.1145/1620545.1620567

[41] A.-S. DADZIE, M. ROWE.*Approaches to Visualising Linked Data: A Survey*, in "Semantic Web", 2011, vol. 2, n^o 2, http://dx.doi.org/10.3233/SW-2011-0037

[42] E. DIMARA, G. BAILLY, A. BEZERIANOS, S. FRANCONERI.*Mitigating the Attraction Effect with Visualizations*, in "IEEE Transactions on Visualization and Computer Graphics", October 2018, vol. 25, n^o 1, p. 850 - 860 [*DOI :* 10.1109/TVCG.2018.2865233], https://hal.inria.fr/hal-01845004

[43] A. DIX, G. LEPOURAS, A. KATIFORI.*From the web of data to a world of action*, in "Web Semantics", 2010, vol. 8, p. 394–408, http://dx.doi.org/10.1016/j.websem.2010.04.007

[44] D. FENSEL, J. HENDLER, H. LIEBERMAN, W. WAHLSTER.*Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, 2005

[45] C. FORLINES, A. ESENTHER, C. SHEN, D. WIGDOR, K. RYALL.*Multi-user, Multi-display Interaction with a Single-user, Single-display Geospatial Application*, in "Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology", UIST '06, ACM, 2006, p. 273–276, http://doi.acm.org/10.1145/1166253.1166296

[46] J. FUCHS, P. ISENBERG, A. BEZERIANOS, F. FISCHER, E. BERTINI.*The Influence of Contour on Similarity Perception of Star Glyphs*, in "IEEE Transactions on Visualization and Computer Graphics", 2014, vol. 20, n^o 12, p. 2251-2260 [*DOI :* 10.1109/TVCG.2014.2346426], https://hal.inria.fr/hal-01023867

[47] J. FUCHS, P. ISENBERG, A. BEZERIANOS, D. KEIM.*A Systematic Review of Experimental Studies on Data Glyphs*, in "IEEE Transactions on Visualization and Computer Graphics", July 2017, vol. 23, n^o 7, p. 1863-1879 [*DOI :* 10.1109/TVCG.2016.2549018], https://hal.inria.fr/hal-01378429

[48] J. GRUDIN.*Computer-Supported Cooperative Work: History and Focus*, in "IEEE Computer", May 1994, vol. 27, n^o 5, p. 19–26, http://dx.doi.org/10.1109/2.291294

[49] T. HEATH, C. BIZER.*Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, 2011

[50] J. HEBELER, M. FISHER, R. BLACE, A. PEREZ-LOPEZ.*Semantic web programming*, Wiley & Sons, 2011

[51] M. HILDEBRAND, J. VAN OSSENBRUGGEN, L. HARDMAN.*/facet: A Browser for Heterogeneous Semantic Web Repositories*, in "ISWC'06: Proc. of the International Semantic Web Conference", Springer, 2006, p. 272–285

[52] D. F. HUYNH, R. C. MILLER, D. KARGER.*Potluck: Data Mash-up Tool for Casual Users*, in "Web Semantics", 2008, vol. 6, n^o 4, p. 274–282, http://dx.doi.org/10.1016/j.websem.2008.09.005

[53] H. ISHII, B. ULLMER.*Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms*, in "Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems", CHI '97, ACM, 1997, p. 234–241, http://doi.acm.org/10.1145/258549.258715

[54] J. A. JACKO.*Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, 3rd Edition*, 3rd, CRC Press, 2012

[55] D. KARGER, S. OSTLER, R. LEE.*The Web Page As a WYSIWYG End-user Customizable Database-backed Information Management Application*, in "Proc. of the Symposium on User Interface Software and Technology", UIST '09, ACM, 2009, p. 257–260, http://doi.acm.org/10.1145/1622176.1622223

[56] A. KATIFORI, C. HALATSIS, G. LEPOURAS, C. VASSILAKIS, E. GIANNOPOULOU.*Ontology visualization methods—a survey*, in "ACM Computing Surveys", 2007, vol. 39, n^o 4, p. 10:1-10:42

[57] C. LIU, O. CHAPUIS, M. BEAUDOUIN-LAFON, E. LECOLINET.*Shared Interaction on a Wall-Sized Display in a Data Manipulation Task*, in "Proceedings of the 34th international conference on Human factors in computing systems", San Jose, United States, ACM (editor), CHI '16, SigCHI and ACM, May 2016, p. 1-12 [*DOI :* 10.1145/2858036.2858039], https://hal.inria.fr/hal-01275535

[58] C. LIU, O. CHAPUIS, M. BEAUDOUIN-LAFON, E. LECOLINET, W. MACKAY.*Effects of Display Size and Navigation Type on a Classification Task*, in "Proceedings of the 32nd international conference on Human factors in computing systems", CHI '14, ACM, 2014, p. 4147-4156, http://doi.acm.org/10.1145/2556288.2557020

[59] W. MACKAY.*Is paper safer? The role of paper flight strips in air traffic control*, in "ACM Trans. Comput.-Hum. Interact.", December 1999, vol. 6, p. 311–340, http://doi.acm.org/10.1145/331490.331491

[60] R. MORALES, C. APPERT, G. BAILLY, E. PIETRIGA.*TouchTokens: Guiding Touch Patterns with Passive Tokens*, in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems", New York, NY, USA, CHI '16, ACM, 2016, To appear

[61] T. MUNZNER.*Visualization Analysis and Design*, CRC Press, 2014

[62] M. NANCEL, O. CHAPUIS, E. PIETRIGA, X.-D. YANG, P. P. IRANI, M. BEAUDOUIN-LAFON.*High-precision pointing on large wall displays using small handheld devices*, in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems", CHI '13, ACM, 2013, p. 831–840, http://doi.acm.org/10.1145/2470654.2470773

[63] M. NANCEL, E. PIETRIGA, O. CHAPUIS, M. BEAUDOUIN-LAFON.*Mid-air Pointing on Ultra-Walls*, in "ACM Transactions on Computer-Human Interaction", August 2015, vol. 22, n° 5, p. 21:1–21:62 [*DOI :* 10.1145/2766448], https://hal.inria.fr/hal-01184544

[64] A.-C. NGONGA NGOMO, L. BÜHMANN, C. UNGER, J. LEHMANN, D. GERBER.*Sorry, I Don'T Speak SPARQL: Translating SPARQL Queries into Natural Language*, in "Proc. of the International Conference on World Wide Web", WWW '13, ACM, 2013, p. 977–988, http://dl.acm.org/citation.cfm?id=2488388.2488473

[65] H. OLAFSDOTTIR, C. APPERT.*Multi-Touch Gestures for Discrete and Continuous Control*, in "International Working Conference on Advanced Visual Interfaces (AVI)", Como, Italy, May 2014, 8 [*DOI :* 10.1145/2598153.2598169], https://hal.archives-ouvertes.fr/hal-00998971

[66] H. OLAFSDOTTIR, T. TSANDILAS, C. APPERT.*Prospective Motor Control on Tabletops: Planning Grasp for Multitouch Interaction*, in "Proceedings of the 32nd international conference on Human factors in computing systems", CHI '14, ACM, 2014, p. 2893-2902, http://doi.acm.org/10.1145/2556288.2557029

[67] E. PIETRIGA, C. APPERT, M. BEAUDOUIN-LAFON.*Pointing and beyond: an operationalization and preliminary evaluation of multi-scale searching*, in "CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems", ACM Press, 2007, p. 1215–1224 [*DOI :* 10.1145/1240624.1240808]

[68] E. PIETRIGA, C. APPERT.*Sigma lenses: focus-context transitions combining space, time and translucence*, in "CHI '08: Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems", ACM, 2008, p. 1343–1352 [*DOI :* 10.1145/1357054.1357264]

[69] E. PIETRIGA, O. BAU, C. APPERT.*Representation-Independent In-Place Magnification with Sigma Lenses*, in "IEEE Transactions on Visualization and Computer Graphics (TVCG)", 2010, vol. 16, n° 03, p. 455-467 [*DOI :* 10.1109/TVCG.2009.98]

[70] E. PIETRIGA, C. BIZER, D. KARGER, R. LEE.*Fresnel - A Browser-Independent Presentation Vocabulary for RDF*, in "Proceedings of the 5th International Semantic Web Conference (ISWC 2006)", Springer, November 2006, p. 158-171

[71] E. PIETRIGA, S. HUOT, M. NANCEL, R. PRIMET.*Rapid development of user interfaces on cluster-driven wall displays with jBricks*, in "Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems", EICS '11, ACM, 2011, p. 185–190, http://doi.acm.org/10.1145/1996461.1996518

[72] E. PIETRIGA.*A Toolkit for Addressing HCI Issues in Visual Language Environments*, in "IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)", IEEE Computer Society, 2005, p. 145-152 [*DOI :* 10.1109/VLHCC.2005.11]

[73] C. PINDAT, E. PIETRIGA, O. CHAPUIS, C. PUECH.*JellyLens: Content-Aware Adaptive Lenses*, in "UIST - 25th Symposium on User Interface Software and Technology - 2012", Cambridge, MA, United States,

Proceedings of the 25th Symposium on User Interface Software and Technology, ACM, October 2012, p. 261-270 [*DOI :* 10.1145/2380116.2380150], https://hal.inria.fr/hal-00721574

[74] C. PINDAT, E. PIETRIGA, O. CHAPUIS, C. PUECH.*Drilling into complex 3D models with gimlenses*, in "Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology", VRST '13, ACM, 2013, p. 223–230, http://doi.acm.org/10.1145/2503713.2503714

[75] H. ROMAT, C. APPERT, B. BACH, N. HENRY-RICHE, E. PIETRIGA.*Animated Edge Textures in Node-Link Diagrams: a Design Space and Initial Evaluation*, in "Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems", Montréal, Canada, CHI '18, ACM, April 2018, p. 187:1–187:13 [*DOI :* 10.1145/3173574.3173761], https://hal.inria.fr/hal-01726358

[76] H. ROMAT.*From data exploration to presentation : designing new systems and interaction techniques to enhance the sense-making process*, Université Paris-Saclay, October 2019, https://tel.archives-ouvertes.fr/tel-02390645

[77] N. SHADBOLT, T. BERNERS-LEE, W. HALL.*The Semantic Web Revisited*, in "IEEE Intelligent Systems", 2006, vol. 21, n^{o} 3, p. 96-101, http://doi.ieeecomputersociety.org/10.1109/MIS.2006.62

[78] O. SHAER, E. HORNECKER.*Tangible User Interfaces: Past, Present, and Future Directions*, in "Found. Trends Hum.-Comput. Interact.", January 2010, vol. 3, n^{o} 1–2, p. 1–137, http://dx.doi.org/10.1561/1100000026

[79] B. SHNEIDERMAN, C. PLAISANT.*Designing the User Interface, 4th Edition*, Pearson Addison Wesley, 2005

[80] D. SPELMEZAN, C. APPERT, O. CHAPUIS, E. PIETRIGA.*Controlling widgets with one power-up button*, in "Proceedings of the 26th annual ACM symposium on User interface software and technology", UIST '13, ACM, 2013, p. 71–74, http://doi.acm.org/10.1145/2501988.2502025

[81] D. SPELMEZAN, C. APPERT, O. CHAPUIS, E. PIETRIGA.*Side pressure for bidirectional navigation on small devices*, in "Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services", MobileHCI '13, ACM, 2013, p. 11–20, http://doi.acm.org/10.1145/2493190.2493199

[82] G. TOUYA, J.-F. GIRRES.*ScaleMaster 2.0: a ScaleMaster extension to monitor automatic multi-scales generalizations*, in "Cartography and Geographic Information Science", 2013, vol. 40, n^{o} 3, p. 192–200, http://dx.doi.org/10.1080/15230406.2013.809233

[83] G. TOUYA.*Social Welfare to Assess the Global Legibility of a Generalized Map*, in "International Conference on Geographic Information", Springer, 2012, p. 198–211, http://dx.doi.org/10.1007/978-3-642-33024-7_15

[84] T. TSANDILAS, C. APPERT, A. BEZERIANOS, D. BONNET.*Coordination of Tilt and Touch in One- and Two-Handed Use*, in "Proceedings of the 32nd international conference on Human factors in computing systems", CHI '14, ACM, 2014, p. 2001-2004, http://doi.acm.org/10.1145/2556288.2557088

[85] T. TSANDILAS, A. BEZERIANOS, T. JACOB.*SketchSliders: Sketching Widgets for Visual Exploration on Wall Displays*, in "Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems", Seoul, South Korea, ACM, April 2015, p. 3255-3264 [*DOI :* 10.1145/2702123.2702129], https://hal.archives-ouvertes.fr/hal-01144312

[86] G. TUMMARELLO, R. DELBRU, E. OREN.*Sindice.com: Weaving the Open Linked Data*, in "Proc. of the 6th Int. Semantic Web Conf. (ISWC2007)", Springer-Verlag, 2007, p. 552–565

[87] C. WARE.*Information visualization: perception for design*, Elsevier, 2012

[88] S. ZHAI, P. O. KRISTENSSON, C. APPERT, T. H. ANDERSEN, X. CAO.*Foundational Issues in Touch-Screen Stroke Gesture Design - An Integrative Review*, in "Foundations and Trends in Human-Computer Interaction", December 2012, vol. 5, n^o^ 2, p. 97-205 [*DOI :* 10.1561/1100000012], https://hal.inria.fr/hal-00765046

[89] R. A. DE ALMEIDA, C. PILLIAS, E. PIETRIGA, P. CUBAUD.*Looking behind bezels: french windows for wall displays*, in "Proceedings of the International Working Conference on Advanced Visual Interfaces", AVI '12, ACM, 2012, p. 124–131, http://doi.acm.org/10.1145/2254556.2254581

Project-Team LIFEWARE

Computational systems biology and optimization

RESEARCH CENTER
Saclay - Île-de-France

THEME
Computational Biology

Table of contents

Project-Team LIFEWARE

Creation of the Team: 2014 January 01, updated into Project-Team: 2015 April 01

Keywords:

Computer Science and Digital Science:

A2.1.1. - Semantics of programming languages
A2.1.5. - Constraint programming
A2.1.10. - Domain-specific languages
A2.2.1. - Static analysis
A2.3.2. - Cyber-physical systems
A2.4. - Formal method for verification, reliability, certification
A2.4.1. - Analysis
A2.4.2. - Model-checking
A2.4.3. - Proofs
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A6.1.4. - Multiscale modeling
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A6.3.1. - Inverse problems
A6.3.4. - Model reduction
A7.2. - Logic in Computer Science
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.7. - Graph theory
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1.2. - Molecular and cellular biology
B1.1.7. - Bioinformatics
B1.1.8. - Mathematical biology
B1.1.10. - Systems and synthetic biology
B2. - Health
B2.2.3. - Cancer
B2.4.1. - Pharmaco kinetics and dynamics
B2.4.2. - Drug resistance
B9. - Society and Knowledge

1. Team, Visitors, External Collaborators

Research Scientists

François Fages [Team leader, Inria, Senior Researcher, HDR]
Grégory Batt [Inria, Senior Researcher, HDR]
Jakob Ruess [Inria, Researcher]
Sylvain Soliman [Inria, Researcher, HDR]

External Collaborator

Denis Thieffry [École Normale Supérieure de Paris, HDR]

Technical Staff

François Bertaux [Institut Pasteur, Engineer]
David Coudrin [Inria, Engineer, until Jul 2019]
Steven Fletcher [Inria, Engineer]
Andjela Davidovic [Institut Pasteur, Engineer, part time]

PhD Students

Chetan Aditya [Inria, PhD Student]
Virgile Andréani [École polytechnique and Inria, PhD Student]
Eléonore Bellot [Inria, PhD Student]
Arthur Carcano [Université Paris Diderot, PhD Student]
Elisabeth Degrand [Inria, PhD Student, from Oct 2019]
Eléa Greugny [Johnson&Johnson France, PhD Student, from Aug 2019, granted by CIFRE]
Jeremy Grignard [Institut de recherche Servier, PhD Student, from Mar 2019]
Julien Martinelli [INSERM, PhD Student]
Sebastian Ramon Sosa Carrillo [Inria, PhD Student]
Elise Weill Duflos [Inria, PhD Student, until Oct 2019, with EPI Commands]

Post-Doctoral Fellows

Olivier Borkowski [Inria, Post-Doctoral Fellow, from Oct 2019]
Zachary Fox [Inria, Post-Doctoral Fellow, from Jul 2019]
Mathieu Hemery [Inria, Post-Doctoral Fellow]
Davin Lunz [Inria, Post-Doctoral Fellow, from Nov 2019, with EPI Commands]

Visiting Scientists

Claudia Lopez Zazueta [NTNU, Norway, Oct 2019]
Lucia Nasti [Univ. Pisa, Italy, until Mar 2019]

Administrative Assistants

Natalia Alves [Inria, Administrative Assistant]
Adeline Lochet [Inria, InBio Administrative Assistant, until May 2019]
Emmanuelle Perrot [Inria, InBio Administrative Assistant, from May 2019]

2. Overall Objectives

2.1. Overall Objectives

This project aims at developing formal methods and experimental settings for understanding the cell machinery and establishing computational paradigms in cell biology. It is based on the vision of **cells as machines, biochemical reaction networks as programs**, and on the use of concepts and tools from computer science to master the complexity of cell processes.

This project addresses fundamental research issues in computer science on the **interplay between structure and dynamics** in large interaction networks, and on **mixed analog-discrete computation**. We contribute to the theory of biochemical computation, and develop since 2002 a modelling, analysis and synthesis software, the Biochemical Abstract Machine, BIOCHAM. The reaction rule-based language of this system allows us to reason about biochemical reaction networks at different levels of abstraction, in the stochastic, differential, discrete, Boolean and hybrid semantics of reaction networks. We develop a variety of static analysis methods before going to simulations and dynamical analyses. We use **quantitative temporal logics** as a mean to formalise biological behaviours with imprecise data and to constrain model building or network synthesis.

A **tight integration between dry lab and wet lab efforts** is also essential for the success of the project. This is achieved through tight collaborations with biologists and experimentalists. Furthermore, half of Lifeware is in the InBio group at Institut Pasteur headed by Grégory Batt who develops an experimental platform for the closed-loop control of intracellular processes. This platform combines hardware (microfluidic device and microscope), software (cell tracking and model-based predictive control algorithms) and liveware (genetically modified living cells). The originality of this project thus also deals with the recourse to advanced microscopy and synthetic biology technologies to perform accurate observations, modifications and **real-time control** at both **single cell and cell population levels**.

Because of the importance of optimization techniques in our research, we keep some activity purely dedicated to optimization problems, in particular on constraint programming methods for computing with partial information systems and solving NP-hard static analysis problems, and on continuous optimization methods for dealing with continuous parameters.

3. Research Program

3.1. Computational Systems Biology

Bridging the gap between the complexity of biological systems and our capacity to model and **quantitatively predict system behaviors** is a central challenge in systems biology. We believe that a deeper understanding of the concept and theory of biochemical computation is necessary to tackle that challenge. Progress in the theory is necessary for scaling, and enabling the application of static analysis, module identification and decomposition, model reductions, parameter search, and model inference methods to large biochemical reaction systems. A measure of success on this route will be the production of better computational modeling tools for elucidating the complex dynamics of natural biological processes, designing synthetic biological circuits and biosensors, developing novel therapy strategies, and optimizing patient-tailored therapeutics.

Progress on the **coupling of models to data** is also necessary. Our approach based on quantitative temporal logics provides a powerful framework for formalizing experimental observations and using them as formal specification in model building. Key to success is a tight integration between *in vivo* and *in silico* work, and on the mixing of dry and wet experiments, enabled by novel biotechnologies. In particular, the use of microfluidic devices makes it possible to measure behaviors at both single-cell and cell population levels *in vivo*, provided innovative modeling, analysis and control methods are deployed *in silico*.

In synthetic biology, while the construction of simple intracellular circuits has shown feasible, the design of larger, **multicellular systems** is a major open issue. In engineered tissues for example, the behavior results from the subtle interplay between intracellular processes (signal transduction, gene expression) and intercellular processes (contact inhibition, gradient of diffusible molecule), and the question is how should cells be genetically modified such that the desired behavior robustly emerges from cell interactions.

3.2. Chemical Reaction Network (CRN) Theory

Feinberg's chemical reaction network theory and Thomas's influence network analyses provide sufficient and/or necessary structural conditions for the existence of multiple steady states and oscillations in regulatory networks. Those conditions can be verified by static analyzers without knowing kinetic parameter values nor making any simulation. In this domain, most of our work consists in analyzing the interplay between the **structure** (Petri net properties, influence graph, subgraph epimorphisms) and the **dynamics** (Boolean, CTMC, ODE, time scale separations) of biochemical reaction systems. In particular, our study of influence graphs of reaction systems, our generalization of Thomas' conditions of multi-stationarity and Soulé's proof to reaction systems [0], the inference of reaction systems from ODEs [0], the computation of structural invariants by constraint programming techniques, and the analysis of model reductions by subgraph epimorphisms now provide solid ground for developing static analyzers, using them on a large scale in systems biology, and elucidating modules.

3.3. Logical Paradigm for Systems Biology

Our group was among the first ones in 2002 to apply **model-checking** methods to systems biology in order to reason on large molecular interaction networks, such as Kohn's map of the mammalian cell cycle (800 reactions over 500 molecules) [0]. The logical paradigm for systems biology that we have subsequently developed for quantitative models can be summarized by the following identifications :

$$\text{biological model = transition system } K$$

$$\text{dynamical behavior specification = temporal logic formula } \phi$$

$$\text{model validation = model-checking } \; K,\ s \models ?\ \phi$$

$$\text{model reduction = sub-model-checking, } K' \subset K \text{ s.t. } K'?,\ s \models \phi$$

$$\text{model prediction = formula enumeration, } \phi \text{ s.t. } K,\ s \models \phi?$$

$$\text{static experiment design = symbolic model-checking, state } s \text{ s.t. } K,\ s? \models \phi$$

$$\text{model synthesis = constraint solving } \; K?,\ s \models \phi$$

$$\text{dynamic experiment design = constraint solving } \; K?,\ s? \models \phi$$

In particular, the definition of a continuous satisfaction degree for **first-order temporal logic** formulae with constraints over the reals, was the key to generalize this approach to quantitative models, opening up the field of model-checking to model optimization [0] This line of research continues with the development of temporal logic patterns with efficient constraint solvers and their generalization to handle stochastic effects.

3.4. Computer-Aided Design of CRNs for Synthetic Biology

The continuous nature of many protein interactions leads us to consider models of analog computation, and in particular, the recent results in the theory of analog computability and complexity obtained by Amaury Pouly [0] and Olivier Bournez, establish fundamental links with digital computation. In a paper published last year [0] We have derived from these results the Turing completeness result of elementary CRNs (without polymerization) under the differential semantics, closing a long-standing open problem in CRN theory. The proof of this result shows how computable function over the reals, described by Ordinary Differential

[0] Sylvain Soliman. A stronger necessary condition for the multistationarity of chemical reaction networks. Bulletin of Mathematical Biology, 75(11):2289–2303, 2013.

[0] François Fages, Steven Gay, Sylvain Soliman. Inferring reaction systems from ordinary differential equations. Journal of Theoretical Computer Science (TCS), Elsevier, 2015, 599, pp.64–78.

[0] N. Chabrier-Rivier, M. Chiaverini, V. Danos, F. Fages, V. Schächter. Modeling and querying biochemical interaction networks. Theoretical Computer Science, 325(1):25–44, 2004.

[0] On a continuous degree of satisfaction of temporal logic formulae with applications to systems biology A. Rizk, G. Batt, F. Fages, S. Soliman International Conference on Computational Methods in Systems Biology, 251-268

[0] Amaury Pouly, "Continuous models of computation: from computability to complexity", PhD Thesis, Ecole Polytechnique, Nov. 2015.

[0] Fages, François, Le Guludec, Guillaume and Bournez, Olivier, Pouly, Amaury. Strong Turing Completeness of Continuous Chemical Reaction Networks and Compilation of Mixed Analog-Digital Programs. In CMSB'17: Proceedings of the fiveteen international conference on Computational Methods in Systems Biology, pages 108–127, volume 10545 of Lecture Notes in Computer Science. Springer-Verlag, 2017.

Equations, namely by Polynomial Initial Value Problems (PIVP), can be compiled into elementary biochemical reactions, furthermore with a notion of analog computation complexity defined as the length of the trajectory to reach a given precision on the result. This opens a whole research avenue to analyze biochemical circuits in Systems Biology, transform behavioural specifications into biochemical reactions for Synthetic Biology, and compare artificial circuits with natural circuits acquired through evolution, from the novel point of view of analog computation and complexity.

3.5. Modeling of Phenotypic Heterogeneity in Cellular Processes

Since nearly two decades, a significant interest has grown for getting a quantitative understanding of the functioning of biological systems at the cellular level. Given their complexity, proposing a model accounting for the observed cell responses, or better, predicting novel behaviors, is now regarded as an essential step to validate a proposed mechanism in systems biology. Moreover, the constant improvement of stimulation and observation tools creates a strong push for the development of methods that provide predictions that are increasingly precise (single cell precision) and robust (complex stimulation profiles).

It is now fully apparent that cells do not respond identically to a same stimulation, even when they are all genetically-identical. This phenotypic heterogeneity plays a significant role in a number of problems ranging from cell resistance to anticancer drug treatments to stress adaptation and bet hedging.

Dedicated modeling frameworks, notably **stochastic** modeling frameworks, such as chemical master equations, and **statistic** modeling frameworks, such as ensemble models, are then needed to capture biological variability.

Appropriate mathematical and computational tools should then be employed for the analysis of these models and their calibration to experimental data. One can notably mention **global optimization** tools to search for appropriate parameters within large spaces, **moment closure** approaches to efficiently approximate stochastic models [0], and (stochastic approximations of) the **expectation maximization** algorithm for the identification of mixed-effects models [0].

3.6. External Control of Cell Processes

External control has been employed since many years to regulate culture growth and other physiological properties. Recently, taking inspiration from developments in synthetic biology, closed loop control has been applied to the regulation of intracellular processes. Such approaches offer unprecedented opportunities to investigate how a cell process dynamical information by maintaining it around specific operating points or driving it out of its standard operating conditions. They can also be used to complement and help the development of synthetic biology through the creation of hybrid systems resulting from the interconnection of in vivo and in silico computing devices.

In collaboration with Pascal Hersen (CNRS MSC lab), we developed a platform for gene expression control that enables to control protein concentrations in yeast cells. This platform integrates microfluidic devices enabling long-term observation and rapid change of the cells environment, microscopy for single cell measurements, and software for real-time signal quantification and model based control. We demonstrated in 2012 that this platform enables controlling the level of a fluorescent protein in cells with unprecedented accuracy and for many cell generations [0].

[0] Moment-based inference predicts bimodality in transient gene expression, C. Zechner C, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, Proceedings of the National Academy of Sciences USA, 9(5):109(21):8340-5, 2012

[0] What population reveals about individual cell identity: estimation of single-cell models of gene expression in yeast, A. Llamosi, A.M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt, PLoS Computational Biology, 9(5): e1003056, 2015

[0] Jannis Uhlendorf, Agnés Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Grégory Batt, Pascal Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. Proceedings of the National Academy of Sciences USA, 109(35):14271–14276, 2012.

More recently, motivated by an analogy with a benchmark control problem, the stabilization of an inverted pendulum, we investigated the possibility to balance a genetic toggle switch in the vicinity of its unstable equilibrium configuration. We searched for solutions to balance an individual cell and even an entire population of heterogeneous cells, each harboring a toggle switch [0].

Independently, in collaboration with colleagues from IST Austria, we investigated the problem of controlling cells, one at a time, by constructing an integrated optogenetic-enabled microscopy platform. It enables experiments that bridge individual and population behaviors. We demonstrated: (i) population structuring by independent closed-loop control of gene expression in many individual cells, (ii) cell–cell variation control during antibiotic perturbation, (iii) hybrid bio-digital circuits in single cells, and freely specifiable digital communication between individual bacteria [0].

3.7. Constraint Solving and Optimization

Constraint solving and optimization methods are important in our research. On the one hand, static analysis of biochemical reaction networks involves solving hard combinatorial optimization problems, for which **constraint programming** techniques have shown particularly successful, often beating dedicated algorithms and allowing to solve large instances from model repositories. On the other hand, parameter search and model calibration problems involve similarly solving hard continuous optimization problems, for which **evolutionary algorithms**, and especially the covariance matrix evolution strategy (CMA-ES) [0] have been shown to provide best results in our context, for up to 100 parameters. This has been instrumental in building challenging quantitative models, gaining model-based insights, revisiting admitted assumptions, and contributing to biological knowledge [00].

4. Application Domains

4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Furthermore, one important application of our research is the development of a **modeling software** for computational systems biology.

4.2. Modeling software for systems biology and synthetic biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (BIOCHAM), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values

[0]Jean-Baptiste Lugagne, Sebastian Sosa Carrillo and Melanie Kirch, Agnes Köhler, Gregory Batt and Pascal Hersen. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. Nature Communications, 8(1):1671, 2017.

[0]Remy Chait, Jakob Ruess, Tobias Bergmiller and Gavsper Tkavcik, Cvalin Guet. Shaping bacterial population behavior through computer-interfaced control of individual cells. Nature Communications, 8(1):1535, 2017.

[0]N. Hansen, A. Ostermeier (2001). Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation, 9(2) pp. 159–195.

[0]Domitille Heitzler, Guillaume Durand, Nathalie Gallay, Aurélien Rizk, Seungkirl Ahn, Jihee Kim, Jonathan D. Violin, Laurence Dupuy, Christophe Gauthier, Vincent Piketty, Pascale Crépieux, Anne Poupon, Frédérique Clément, François Fages, Robert J. Lefkowitz, Eric Reiter. Competing G protein-coupled receptor kinases balance G protein and β-arrestin signaling. Molecular Systems Biology, 8(590), 2012.

[0]Pauline Traynard, Céline Feillet, Sylvain Soliman, Franck Delaunay, François Fages. Model-based Investigation of the Circadian Clock and Cell Cycle Coupling in Mouse Embryonic Fibroblasts: Prediction of RevErb-alpha Up-Regulation during Mitosis. Biosystems, 149:59–69, 2016.

satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

4.3. Coupled models of the cell cycle and the circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characterictics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR HYCLOCK project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio C5SYS and FP6 TEMPO projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints [0]. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

4.4. Biosensor design and implementation in non-living protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of high-level functions in non-living vesicles for medical applications, such as biosensors for medical diagnosis [0]. Our approach is based on purely protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sys2Diag which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis, in contrast to living cells in synthetic biology, is particularly appealing for security considerations and compliance to forthcoming EU regulation.

4.5. Functional characterization of the resistance of bacterial populations to antimicrobial treatments

Antibiotic resistance is becoming a problem of central importance at a global level. Two mechanisms are at the origin of non-susceptibility to antimicrobial treatments. The first one comes from adaptation of bacterial cells to antibacterial treatments, notably through the modification of efflux pumps or the expression of enzymes that degrade the antibiotics. Cells are individually resistant. The second one, typically found in resistances to β-lactams, a broad class of antibiotics, originates from the release in the environment of the antibiotic degrading enzymes by the dead cells. This leads to population effects by which cells become collectively resilient.

The functional characterization of these different effects is important for the best use of antibiotics (antibiotic stewardship). In collaboration with Lingchong You (Duke University) and with Philippe Glaser (Institut Pasteur), we develop experimental platforms, models, and optimal model calibration methods that gives precise estimations of individual resistance and collective resilience of bacterial populations to antibiotic treatments.

[0] Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. Theoretical Computer Science, 412(21):2108 2127, 2011.

[0] Alexis Courbet, Patrick Amar, François Fages, Eric Renard, Franck Molina. Computer-aided biochemical programming of synthetic microreactors as diagnostic devices. Molecular Systems Biology, 14(4), 2018.

5. Highlights of the Year

5.1. Highlights of the Year

Figure 1. La Recherche magazine award 2019 ceremony.

- **Creation of a new team at Inria Paris**

  At the end of 2019, the Lifeware team gave birth to a new Inria team, called InBio and affiliated to the Inria Paris research centre. So far, InBio was a Pasteur research unit that was hosting a fraction of the members of the Lifeware team on the campus of Institut Pasteur. So in 2020, InBio becomes a new Common Project-Team between Inria Paris and Pasteur Institute. This demonstrates that Inria is actively supporting research in the computational systems biology field.

- **Launching of Inria Exploratory Action GRAM on chemical programming of artificial vesicles**

  Chemical reaction networks are a computation paradigm used by natural cells to process information, take decisions and control their vital processes. The synthesis of artificial vesicles without DNA nor RNA but containing precise quantities of enzymes allows us today to implement high-level functions in proto-cells with numerous potential applications in health and the environment. Based on previous work of the Lifeware project-team on chemical analog computation theory and programming, of the CNRS-Alcediag Sys2diag laboratory on the synthesis of biosensors in artificial vesicles, and on the expertise of the Roscoff Biological Station on membrane transporters, we explore an original approach to analog chemical circuit design applied to the programming of high-level functions in chemical analog computers.

5.1.1. Awards

- **Award Ceremony - La Recherche magazine 2019 - Information Sciences**

  The ceremony for awards La Recherche magazine 2019 at University Paris-Dauphine was a great occasion to present our article "Strong Turing Completeness of Continuous Chemical Reaction Networks and Compilation of Mixed Analog-Digital Programs" by F. Fages, G. Le Guludec, O. Bournez and A. Pouly, Best Paper award at CMSB 2017, recipient of La Recherche magazine 2019 Award - Information Sciences.

6. New Software and Platforms

6.1. BIOCHAM

The Biochemical Abstract Machine

KEYWORDS: Systems Biology - Bioinformatics

FUNCTIONAL DESCRIPTION: The Biochemical Abstract Machine (BIOCHAM) is a software environment for modeling, analyzing and synthesizing biochemical reaction networks (CRNs) with respect to a formal specification of the observed or desired behavior of a biochemical system. BIOCHAM is compatible with the Systems Biology Markup Language (SBML) and contains some unique features about formal specifications in quantitative temporal logic, sensitivity and robustness analyses and parameter search in high dimension w.r.t. behavioral specifications, static analyses, and synthesis of CRNs.

RELEASE FUNCTIONAL DESCRIPTION: – notebooks of Master classes – graphical user interface on top of Jupyter – synthesis of at most binary reactions with minimisation of the variables for negative values – detection of model reduction by subgraph epimorphism – option for partial tropical equilibrations

- Participants: François Fages, David Coudrin, Sylvain Soliman and Mathieu Hemery
- Contact: François Fages
- URL: http://lifeware.inria.fr/biocham4/

6.2. casq

CellDesigner as SBML-Qual

KEYWORDS: SBML - Logical Framework - Knowledge representation

FUNCTIONAL DESCRIPTION: CaSQ transforms a big knowledge map encoded as an SBGN-compliant SBML file in CellDesigner into an executable Logical Model in SBML-Qual

- Authors: Sylvain Soliman and Anna Niarakis
- Partner: Université d'Evry-Val d'Essonne
- Contact: Sylvain Soliman
- URL: https://gitlab.inria.fr/soliman/casq/

6.3. Platforms

6.3.1. Smart experimental platforms to automate microbiology experiments

Models play a central role in our work, either to test our understanding or to guide the design of novel systems. Model development and parameter calibration necessitate informative experiments. We develop methods to assist with the optimal design of experiments. In consequence, we have to perform, in sequence or in parallel, experiments with possibly complex input profiles. This led us to develop experimental platforms that allow for flexible and automated stimulations and measurements. Three platforms are being developed, based on (i) a microplate photometer, (ii) a bioreactor platform coupled with a flow cytometer, and (iii) a microscope equipped with microfluidic systems, respectively. In all cases, the real-time measurement and actuation capabilities allow for making reactive experiments, notably including real-time control experiments.

7. New Results

7.1. CRN design by program compilation

Participants: Elisabeth Degrand, François Fages, Mathieu Hemery, Wei-Chih Huang [NTU Taiwan], Sylvain Soliman.

One goal of synthetic biology is to implement useful functions with biochemical reactions, either by reprogramming living cells or programming artificial vesicles. In this perspective, we consider Chemical Reaction Networks (CRN) as a programming language, and investigate the CRN program synthesis problem. Recent work has shown that CRN interpreted by differential equations are Turing-complete and can be seen as analog computers where the molecular concentrations play the role of information carriers. Any real function that is computable by a Turing machine in arbitrary precision can thus be computed by a CRN over a finite set of molecular species. The proof of this result gives a numerical method to generate a finite CRN for implementing a real function presented as the solution of a Polynomial Initial Values Problem (PIVP).

The compilation of high-level imperative programming languages in CRN requires however an efficient implementation of program control flows using threshold functions. The biochemical threshold function is also a crucial component in the biosensor circuits to be deployed in living cells or synthetic vesicles for disease diagnosis. In [5], based on the zero-order ultrasensitivity, we propose an economic biochemical implementation of threshold functions with reconfigurable threshold values. We show that the so-constructed threshold function module well approximates the unit step function and allows robust composition with other function modules for complex computation tasks. This is now implemented in BIOCHAM-4 for the compilation of sequentiality and conditionals in CRNs.

7.2. CRN design by artificial evolution

Participants: Elisabeth Degrand, François Fages, Mathieu Hemery, Sylvain Soliman.

In [4], [12], we study an alternative method based on artificial evolution to build a CRN that approximates a real function given on finite sets of input values. We present a nested search algorithm that evolves the structure of the CRN and optimizes the kinetic parameters at each generation. We evaluate this algorithm on the Heaviside and Cosine functions both as functions of time and functions of input molecular species. We then compare the CRN obtained by artificial evolution both to the CRN generated by the numerical method from a PIVP definition of the function, and to the natural CRN found in the BioModels repository for switches and oscillators.

On a Heaviside function of time, the results obtained by artificial evolution lead to a remarkably simple CRN of 3 molecular species and 5 reactions with double catalysts which provide a very stiff transition although using mass action law kinetics. This solution is more economical than the CRN generated by the PIVP method for sigmoid functions. On a Heaviside function of input, the CRN found by evolution are slightly more complicated than the bistable switch found in cell cycle CRN for instance, but much less complex than the MAPK signaling network that plays a similar role.

On the cosine function of time, the best CRN found by evolution contains an annihilation reaction similar to the CRN generated by the numerical method for positive and negative variables, but one less reaction thanks to an intriguing non symmetric use of the two variables which preserves the limit cycle. Interestingly, the evolved and the PIVP generated structures could be compared to prokaryote and eukaryote models of the circadian clock found in BioModels.

On the cosine function of input, a CRN surprisingly emerges with the structure of the CRN for cosine function of time, using the same trick as for PIVP compilation to stop time at the desired input value.

In [2], we use a genetic algorithm to evolve biochemical networks displaying a direct logarithmic response. Numerous biological systems are known to harbour a form of logarithmic behaviour, from Weber's law to bacterial chemotaxis. Working on a logarithmic scale allows the organism to respond appropriately to large variations in a given input at a modest cost in terms of metabolism. Interestingly, a quasi-perfect log-response implemented by the same simple core network evolves in a convergent way across our different replications. The best network is able to fit a logarithm over 4 order of magnitude with an accuracy of the order of 1%. At the heart of this network, we show that a logarithmic approximation may be implemented with one single non-linear interaction, that can be interpreted either as a phosphorylation or as a ligand induced multimerization and provide an analytical explanation of the effect. Biological log-response might thus be easier to implement than usually assumed.

7.3. CRN learning from data time series

Participants: François Fages, Jeremy Grignard, Julien Martinelli, Sylvain Soliman.

With the automation of biological experiments and the increase of quality of single cell data that can now be obtained by phosphoproteomic and time lapse videomicroscopy, automating the building of mechanistic models from these data time series becomes conceivable and a necessity for many new applications. While learning numerical parameters to fit a given model structure to observed data is now a quite well understood subject, learning the structure of the model is a more challenging problem that previous attempts failed to solve without relying quite heavily on prior knowledge about that structure. In [8], [7], we consider mechanistic models based on chemical reaction networks (CRN) with their continuous dynamics based on ordinary differential equations, and finite time series about the time evolution of concentration of molecular species for a given time horizon and a finite set of perturbed initial conditions. We present a greedy heuristics unsupervised statistical learning algorithm to infer reactions with a time complexity for inferring one reaction in $O(t.n^2)$where n is the number of species and t the number of observed transitions in the traces. We evaluate this algorithm both on simulated data from hidden CRNs, and on real videomicroscopy single cell data about the circadian clock and cell cycle progression of NIH3T3 embryonic fibroblasts. In all cases, our algorithm is able to infer meaningful reactions, though generally not a complete set for instance in presence of multiple time scales or highly variable traces.

7.4. CRN reductions

Participants: Orianne Bargain, Eléonore Bellot, François Fages, Eva Philippe, Sylvain Soliman.

We have shown in the past that model reduction relationships between CRNs can be detected on a large scale by the graph matching notion of subgraph epimorphism [0], furthermore quite efficiently using constraint programming or SAT solving techniques. Nevertheless, establishing whether two models are linked through a SEPI is an NP-complete problem which can be computationally expensive in some practical cases. Furthermore, the number of SEPIs can be huge, and some of them may not have a biological interpretation. In [11], we have improved the SEPI framework in this respect in three ways: by introducing optimization criteria to restrict the set of solutions, by restricting merge operations to some notion of neighborhood, and by preprocessing the CRN graphs in normal form in order to eliminate some common model reduction patterns.

Furthermore, in the framework of the ANR-DFG SYMBIONT project we investigate mathematical justification of SEPI reductions based on Tikhonov's theorem and their computation using tropical algebra methods and constraint programming techniques [0].

[0] Steven Gay, François Fages, Thierry Martinez, Sylvain Soliman, Christine Solnon. On the subgraph Epimorphism Problem. Discrete Applied Mathematics, 162:214–228, 2014.

[0] Sylvain Soliman, François Fages, Ovidiu Radulescu. A constraint solving approach to model reduction by tropical equilibration. Algorithms for Molecular Biology, 9(24), 2014.

7.5. CRN modeling of biological systems

Participants: Auriane Cozic, Elisabeth Degrand, François Fages, Eléa Greugny, Jeremy Grignard, Constance Le Gac, Léna Le Quellec, Paul Remondeau, Sylvain Soliman.

This year, beyond implementation work on hybrid simulations in BIOCHAM and on antithetic feedback control in CRNs, we have started the computational modelling of three biological systems with important potential applications in biomedicine.

The first is about erythrocytes (i.e. red blood cells). Their most obvious function concerns the respiratory system since erythrocytes allow gas exchanges at the level of the organism by transporting dioxygen and carbon dioxide between the lungs and the tissues. However, red blood cells also have an important buffer function in the blood, which is necessary to keep blood pH in the physiological range. Modelling the red blood cells with CRNs using BIOCHAM gives us insight as to which biological objects are necessary to allow the cell to process its functions correctly. At the level of Systems Biology, it also allows us to understand the links between the different biological functions of erythrocytes.

The second concerns microtubules and their post-translational modifications involved in major cellular processes such as: mitosis, cardiomyocyte contraction, and neuronal differentiation. More precisely, in neurons, the post-translational modifications of detyrosination and tyrosination are crucial for neuronal plasticity, axon regeneration, recruitment and transports of proteins and correct neuronal wiring. We hypothesize that the decrease of density and length of microtubules and the loss of neuronal structures such as synapses, dendritic spine and growth cone which are correlated with the progressive cognitive decline [9,10] may be the consequence of the dysregulation of the cycle detyrosination/tyrosination in neurodegenerative disorder. This hypothesis is investigated in collaboration with Servier by combining experimental approaches with mathematical modelling.

The third concerns inflammation processes in skin. Skin protects the body against external agents, for instance pathogens, irritants, or UV radiation, that can trigger inflammation. Inflammation is a complex phenomenon that is classified in two main types, acute and chronic. They are distinguished by different parameters such as the duration, the underlying mechanisms, the components involved like the type of immune cells, and the nature and intensity of the associated clinical signs. The computational models developed in collaboration with Johnson&Johnson France, combine mathematical and multi-agent modelling using BIOCHAM and EPISIM modelling tools.

7.6. Automated Inference of Boolean models from molecular interaction maps

Participant: Sylvain Soliman.

Molecular interaction maps have emerged as a meaningful way of representing biological mechanisms in a comprehensive and systematic manner. However, their static nature provides limited insights to the emerging behavior of the described biological system under different conditions. Computational modeling provides the means to study dynamic properties through *in silico* simulations and perturbations.

In collaboration with Anna Niarakis (Université d'Évry, GenHotel) we have started developing the CaSQ Python package, by defining simplification rules and logical formulas for the inferred Boolean models according to the topology and the annotations of the starting molecular interaction maps. We used CaSQ to produce executable files of existing molecular maps notably a big map of the Rheumatoid Arthritis that is at the core of Évry team's work.

A publication on the inference process has already been submitted to Bioinformatics but work continues on the applications side to fine-tune the automatically generated model and analyze its dynamical properties.

7.7. Optimal control of an artificial microbial differentiation system for protein bioproduction

Participants: Élise Weill Duflos, Virgile Andréani, Chetan Aditya, Pierre Martinon [EPI Commands], Jakob Ruess, Grégory Batt, J. Frédéric Bonnans [EPI Commands].

The growth of microorganisms is controlled by strategies for the dynamical allocation of available resources over different cellular functions. Synthetic biology approaches are considered nowadays to artificially modify these strategies and turn microbial populations into biotechnological factories for the production of metabolites of interest. In our recent work, we have studied dynamics of microbial resource allocation and growth in terms of coarse-grained self-replicator models described by ordinary differential equations, and proposed artificial control strategies for the optimization of metabolite production based on the reengineering of resource allocation. In this contribution, we elaborated on our earlier results and further investigate synthetic resource allocation control strategies [9]. Using numerical simulation, we studied the effect on growth and bioproduction of the (biological or technological) costs associated with discontinuous control strategies, and of the time allotted to optimal substrate utilization. Results provided novel insight into the most favorable synthetic control strategies.

7.8. Can optimal experimental design serve as a tool to characterize highly non-linear synthetic circuits?

Participants: Maxim Kryukov [Pasteur Institute], Arthur Carcano, Grégory Batt, Jakob Ruess.

One of the most crippling problems in quantitative and synthetic biology is that models aiming to describe the real mechanisms of biochemical processes inside cells typically contain too many unknown parameters to be reliably inferable from available experimental data. Recent years, however, have seen immense progress in the development of experimental platforms that allow not only to measure biological systems more precisely but also to administer external control inputs to the cells. Optimal experimental design has been identified as a tool that can be used to decide how to best choose these control inputs so as to excite the systems in ways that are particularly useful for learning the biochemical rate constants from the corresponding data. Unfortunately, the experiment that is best to learn the parameters of a system depends on the precise values of these parameters, which are naturally unknown at the time at which experiments need to be designed. Here, we used a recently constructed genetic toggle switch as a case study to investigate how close to the best possible experiment we can hope to get with the most widely used optimal design approaches in the field. We found that, for strongly nonlinear systems such as the toggle switch, reliably predicting the information that can be gained from a priori fixed experiments can be difficult if the system parameters are not known very precisely [6]. This suggests that a better strategy to guarantee informative experiments might be to use feedback control and to adjust the experimental plan in real time.

7.9. Molecular noise of innate immunity shapes bacteria-phage ecologies

Participant: Jakob Ruess.

Mathematical models have been used successfully at diverse scales of biological organization, ranging from ecology and population dynamics to stochastic reaction events occurring between individual molecules in single cells. Generally, many biological processes unfold across multiple scales, with mutations being the best studied example of how stochasticity at the molecular scale can influence outcomes at the population scale. In many other contexts, however, an analogous link between micro-and macro-scale remains elusive, primarily due to the challenges involved in setting up and analyzing multi-scale models. In [3], we employed such a model to investigate how stochasticity propagates from individual biochemical reaction events in the bacterial innate immune system to the ecology of bacteria and bacterial viruses. We showed analytically how the dynamics of bacterial populations are shaped by the activities of immunity-conferring enzymes in single cells and how the ecological consequences imply optimal bacterial defense strategies against viruses. Our results suggest that bacterial populations in the presence of viruses can either optimize their initial growth rate or their population size, with the first strategy favoring simple immunity featuring a single restriction modification system and the second strategy favoring complex bacterial innate immunity featuring several simultaneously active restriction modification systems.

7.10. Estimating information in time-varying signals

Participant: Jakob Ruess.

Across diverse biological systems - ranging from neural networks to intracellular signaling and genetic regulatory networks - the information about changes in the environment is frequently encoded in the full temporal dynamics of the network nodes. A pressing data-analysis challenge has thus been to efficiently estimate the amount of information that these dynamics convey from experimental data. In [1], we developed and evaluated decoding-based estimation methods to lower bound the mutual information about a finite set of inputs, encoded in single-cell high-dimensional time series data. For biological reaction networks governed by the chemical Master equation, we derived model-based information approximations and analytical upper bounds, against which we benchmarked our proposed model-free decoding estimators. In contrast to the frequently-used k-nearest-neighbor estimator, decoding-based estimators robustly extract a large fraction of the available information from high-dimensional trajectories with a realistic number of data samples. We applied these estimators to previously published data on Erk and Ca^{2+} signaling in mammalian cells and to yeast stress-response, and found that substantial amount of information about environmental state can be encoded by non-trivial response statistics even in stationary signals. We argued that these single-cell, decoding-based information estimates, rather than the commonly-used tests for significant differences between selected population response statistics, provide a proper and unbiased measure for the performance of biological signaling networks.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contract with Institut de recherche Servier

In the framework of the Cifre PhD thesis of Jeremy Grignard at Servier, we work on the coupling between computational modeling and biological experiment design, and on chemical reaction network inference methods from data time series.

8.2. Bilateral Grant with Johnson&Johnson France

In the framework of the Cifre PhD thesis of Eléa Greugny at Johnson&Johnson Santé Beauté France, we work on the computational modeling of inflammatory process in the skin, using multi-scale modeling and multi-agent simulation.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR Projects

- ANR-FWF CyberCircuits (2018-2022): “Cybergenetic circuits to test composability of gene networks”, co-coordinated by C. Guet (IST Austria, Klosterneuburg, Austria) and J. Ruess (Inria EPI Lifeware);
- ANR-DFG SYMBIONT (2018-2021) on “Symbolic Methods for Biological Systems”, coordinated by T. Sturm (CNRS, LORIA, Nancy, France) and A. Weber (Univ. Bonn, Germany) with F. Fages and F. Boulier (U. Lille), O. Radulescu (U. Montpellier), A. Schuppert (RWTH Aachen), S. Walcher (RWTH Aachen), W. Seiler (U. Kassel);
- ANR-MOST BIOPSY (2016-2020) on “Biochemical Programming System”, coordinated by F. Molina (CNRS, Sys2diag, Montpellier) and J.H. Jiang (National Taiwan University), with F. Fages;

- ANR MEMIP (2016-2020) on “Mixed-Effects Models of Intracellular Processes”, coordinated by G. Batt, with P. Hersen, (CNRS/Paris7), E. Cinquemani (Inria EPI IBIS) and M. Lavielle (Inria/CNRS/Polytechnique, EPI XPOP);
- ANR COGEX (2016-2019) on “Computer Aided Control of Gene Expression” coordinated by P. Hersen (MSC lab, CNRS/Paris7), with G. Batt and G. Truan (LISBP, CNRS/INSA);

9.1.2. Inria Project Lab

- IPL COSY (2017-2021) on “real-time control of synthetic microbial communities”, coordinated by Eugenio Cinquemani (Ibis, Inria), with Jean-Luc Gouzé (Biocore, Inria), Grégory Batt, Frédéric Bonnans (Commands, Inria), Efimov Denis (Non-A, Inria), and Hans Geiselmann (BIOP, Université Grenoble-Alpes), Béatrice Laroche (Maiage, Inra Jouy-en-Josas).

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

- H2020 FET-OPEN COSY-BIO (2017-2020), on “Control Engineering of Biological Systems for Reliable Synthetic Biology Applications”, coordinated by Diego di Bernardo (Tigem), with Filippo Menolascina (Edinburgh U), Mario di Bernardo (Naples U), Pascal Hersen (Paris7 U), Mustafa Khammash (ETHZ), Grégory Batt, Guy-Bart Stan (Imperial College), and Lucia Marucci (Bristol U).

9.3. International Research Visitors

9.3.1. Visits of International Scientists

The following researchers have been invited for short visits:

- Jean-Louis Lassez, retired IBM Yorktown, USA
- Lucia Nasti, Univ. Pisa, Italy
- Claudia Lopez Zazueta, NTNU, Norway

9.3.1.1. Internships

- Orianne Bargain (TU Dresden Germany)
- Elisabeth Degrand (KTH, Stockholm Sweden)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

Chetan Aditya was the co-organiser of a thematic workshop entitled "From Bench to Bedside" at the Center for Interdisciplinary research, Paris, April 2019.

Grégory Batt was a co-organizer of the thematic session entitled "Predictive approaches for biological systems engineering" at the JOBIM conference, Nantes, July 2019.

François Fages and Sylvain Soliman were scientific co-chairs of

- the Constraints and Life Sciences Track of CP’2019 25th International Conference on Principles and Practice of Constraint Programming, Stamford, CT, USA, September 30 to October 4, 2019.

François Fages was scientific co-chair of

- Workshop on Computational Systems Biology for Complex Disease CSBCD, ENS Paris-Saclay, Cachan, France, 28-29 Nov. 2019.
- Formal methods for the synthesis of biomolecular circuits Shonan Village, Japan, 2-6 Sep 2019.
- France-Taiwan Summer School on New Strategies in Medical Diagnosis and Precision Medicine, NTU, Taipei, Taiwan, 9-10 July 2019.

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- Grégory Batt was a member of the program committee of the IFAC Conference on Foundations of Systems Biology in Engineering (FOSBE 2019), Valencia, Spain, and of the scientific committee of Advanced Lecture Course on Computational Systems Biology summer school, Aussois, France.
- François Fages was PC member of
  - CIBCB'19 16th IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology – Certosa di Pontignano, Siena - Tuscany, Italy, July 9-11, 2019
  - CMSB'19 17th International Conference on Computational Methods in Systems Biology. Trieste, Italy, September 2019.
  - HSB'19, 6th International Workshop on Hybrid Systems and Biology, will be held at the Charles University, Prague (CZ) on the 6th and 7th April 2019, and is colocated with ETAPS 2019.
  - BIOINFORMATICS'19 10th International Conference on Bioinformatics Models, Methods and Algorithms" co-located with BIOSTEC'19, Prague, Czech republic, Feb 2019.
- Jakob Ruess was a PC member of CMSB'19, the 17th International Conference on Computational Methods in Systems Biology. Trieste, Italy, September 2019.
- Sylvain Soliman was a PC member of CSBio 2019 10th International Conference on Computational Systems-Biology and Bioinformatics — Nice (France), December 4–7, 2019

10.1.2.2. Reviewer

Jakob Ruess was a reviewer for two conferences: *European Control Conference (ECC)* and *Bioinformatics 2020*.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

François Fages is member of

- the Editorial Board of the Computer Science area of the Royal Society Open Science journal, since 2014;
- the Editorial Board of the journal RAIRO OR Operations Research, since 2004.

10.1.3.2. Reviewer - Reviewing Activities

Grégory Batt was a reviewer for *Bioinformatics* and *ACS Synthetic Biology* journals.

François Fages made reviews for the journals *Biosystems* and *Transactions on Computational Biology and Bioinformatics*.

Mathieu Hemery made 5 reviews for the journal *Physical Review E*, and was a reviewer for the journals *Physical Review Letters* and *Royal Society Open Science*.

Jakob Ruess was a reviewer for the journals *Entropy* and *Scientific Reports*.

Sylvain Soliman was a reviewer for *Briefings in Bioinformatics*, and for the *Journal of Theoretical Biology*.

10.1.4. Invited Talks

- Chetan Aditya gave a presentation entitled "Synthetic optogenetic differentiation system for bioproduction in budding yeast" at the EMBL workshop "Creating is Understanding: Synthetic Biology Masters Complexity", September 2019, Heidelberg, Germany
- Grégory Batt gave the following invited talks:
  - Driving cellular processes with quantitative accuracy using real-time control approaches, 6th Cross Disciplinary Genomics Meeting of Sorbonne University, Nov 2019, Paris, France
  - Driving cellular processes with quantitative accuracy using real-time control approaches, SynGen Series UK, Nov 2019, London, UK
  - An experimental platform for automated calibration of antimicrobial resistance models, APHP/IP AMR STORM, Oct 2019, Paris, France
  - Balancing a genetic toggle switch by real-time control and periodic stimulations, International Workshop on Control Engineering and Synthetic Biology, Sept 2019, Oxford, UK
  - Experimental and computational methods for modeling cellular processes, Shonan meeting on Formal methods for the synthesis of biomolecular circuits, Sept 2019, Shonan, Japan
  - Driving cellular processes with quantitative accuracy using real-time control approaches, Seminar of Unité Physico Chimie Curie, Jan 2019, Paris, France
- François Bertaux gave an invited talk at the JOBIM thematic session entitled "Methods for single-cell omics data analysis", Nantes, July 2019.
- Andjela Davidovic gave a presentation entitled "Calibration of stochastic biochemical models using single-cell video-microscopy experiments" at the Stochastic models for biology conference (BioHasard 2019), August 2019, Rennes, France
- François Fages gave invited talk to
  - Turing-completeness of Chemical Reaction Networks in Natural cells and Artificial vesicles (keynote CSBio 2019) The 10th International Conference on Computational Systems-Biology and Bioinformatics. Nice, France. December 4 to 7, 2019
  - Cells as Analog Chemical Computers: Turing Completeness and Synthesis Formal methods for the synthesis of biomolecular circuits Shonan Village, Japan, 2-6 Sep 2019.
  - Modeling and Design of Biological Systems, France-Taiwan Summer School on New Strategies in Medical Diagnosis and Precision Medicin, NTU, Taipei, Taiwan, 9-10 July 2019.
  - La Cellule un Calculateur Chimique Journées scientifiques Inria, Lyon, 5-7 juin 2019.
  - La Cellule un Calculateur Chimique Dassault-Systèmes, Vélizy, June 4th 2019.
  - Calculs Analogiques dans les Programmes Biochimiques Naturels et Synthétiques Colloque d'ouverture 50 ans du Laboratoire Jacques-Louis Lions, Roscoff, France, 4-8 mars 2019.
  - Calculs analogiques dans les réseaux biochimiques naturels et synthétiques Prix La recherche - 15th edition. Université Paris Dauphine, 4 Feb. 2019.
- Jakob Ruess gave an invited talk at the SYMBIONT meeting entitled "Molecular noise shapes bacteria-phage ecologies", Paris, April 2019.

10.1.5. Leadership within the Scientific Community

- Grégory Batt is co-animator of the working group on Symbolic Systems Biology (GT Bioss).
- Grégory Batt is a member of
  - the Technical Committee on Systems Biology of IEEE and CSS societies

   - the scientific board of the French research network on Bioinformatics (GdR BIM)
   - the scientific committee of the Advanced Course on Computational Systems Biology summer school, in Aussois
- François Fages is member of
   - the Steering Committee of the International Conference on Computational Methods in Systems Biology, CMSB, since 2008.
   - the Scientific Committee of the Doctorate School ED 474 FIRE, ex Frontiers in Life Sciences, FdV
   - the Scientific Committee of the Summer School Modélisation Formelle de Réseaux de Régulation Biologique Ile de Porquerolles du 23 au 28 juin 2019 .

10.1.6. Scientific Expertise

François Fages was

- member of the Jury for Inria Awards 2019
- evaluator of two DFG grant proposals
- reviewer of a program proposal from Institut Pascal - Paris Saclay
- reviewer of professorship application, Indraprastha Institute of Information Technology - Delhi, India.

10.1.7. Research Administration

- Grégory Batt is the deputy director of the department of Computational Biology at Institut Pasteur
- François Fages is member of the “Comité des Projets du centre” Inria Saclay-IdF, and Inria representative for Doctorate School Institut Polytechnique de Paris
- Sylvain Soliman is member of the “Commission Scientifique” of Inria Saclay-IdF

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Summer school: François Fages (teacher 6h) Modélisation Formelle de Réseaux de Régulation Biologique, Ile de Porquerolles.

Summer school: Jakob Ruess (teacher 6h) Blackboard course on Optimally learning dynamical models from data at the "Advanced lecture course on computational systems biology" in Aussois, France, April 2019.

Master: Grégory Batt (coordinator and teacher: 35h) and Jakob Ruess (25h), *Computational Biology*, M1, Interdisciplinary Approaches to Life Sciences (AIRE-LiSc).

Master: Grégory Batt (3h) *Synthetic Biology and Control course in Molecular and Cellular Biology* Sorbonne Université, Paris.

Master: François Bertaux (co-coordinator and teacher: 20h), *Systems Biology*, M1, Master Interdisciplinary Approaches to Life Sciences (AIRE-LiSc).

Master: François Fages (coordinator module 24h and teacher 12h) *C2-19 Biochemical Programming*, Master Parisien de Recherche en Informatique (MPRI), Paris.

Master: François Fages (co-coordinator module 36h and teacher 18h) and Sylvain Soliman (co-coordinator, teacher 18h) INF555 - Constraint-based Modeling and Algorithms for Decision Making ProblemsMaster Artificial Intelligence, Ecole Polytechnique.

Bachelor 3: François Fages (co-coordinator module and teacher 4h30) and Sylvain Soliman (co-coordinator module, teacher 4h30) and Mathieu Hemery (TD 10h) CSE 301b: Constraint Logic Programming.

Bachelor 2: Eléonore Bellot (teacher 64h) *CSE201 Object-oriented Programming in C++* TD and project supervision

L1: Julien Martinelli (teacher 64h) Mathematics and Calculus, Univ. Paris Descartes

10.2.2. Supervision

PhD in progress: Chetan Aditya, "Control of heterogenous synthetic microbial systems", ED FdV, Université Sorbonne Paris Cité, Feb. 2018, Grégory Batt

PhD in progress: Virgile Andréani, Calibration efficace de modèles de résistance bactérienne aux antibiotiques à l'aide d'un plan d'expériences optimal, ED IPP, Ecole Polytechnique, Sept. 2016, Grégory Batt

PhD in progress : Eléonore Bellot, "Réduction de modèles différentiels par résolution de contraintes d'algèbre tropicale (min,+)", ED IPP, Ecole Polytechnique, Sept. 2018, F. Fages & S. Soliman (50-50%)

PhD in progress: Arthur Carcano, "Iterative design of single-cell experiments to learn single-cell models of biological systems", ED FdV, Université Sorbonne Paris Cité, Oct. 2018, Jakob Ruess and Grégory Batt

PhD in progress : Elisabeth Degrand, "Chemical Programming in Non-living Vesicles", ED IPP, Ecole Polytechnique, Oct. 2019, F. Fages & S. Soliman (50-50%)

PhD in progress : Jérémy Grignard, "Apprentissage de modèles à partir de données pour la conception d'expériences de criblage et la recherche de médicaments", ED IPP, Ecole Polytechnique, dec. 2018, F. Fages & T. Dorval, Servier (50-50%)

PhD in progress : Eléa Greugny, "Development and Implementation of a Mathematical Model of Inflammation in the Human Skin", ED IPP, Ecole Polytechnique, Aug. 2019, F. Fages & J. Bensaci & G. Stamatas, Johnson&Johnson Santé Beauté France (1/3-2/3))

PhD in progress : Julien Martinelli, "Apprentissage de modèles mécanistes à partir de données temporelles, application à la personnalisation de la chronothérapie des cancers", ED IPP, Ecole Polytechnique, Oct. 2018, F. Fages & A. Ballesta, Inserm (50-50%)

PhD in progress: Sebastian Sosa Carrillo, "Understanding the cost of protein production in yeast", ED FdV, Université Sorbonne Paris Cité, Feb. 2018, Grégory Batt

Master's Thesis: Orianne Bargain, "Graph matching, theory and SAT implementation", TU Dresden, Germany, Sep. 2019, F. Fages & S. Soliman (50-50%)

Master's Thesis: Elisabeth Degrand, "Evolving Chemical Reaction Networks", KTH Stockholm, Sweden, June 2019, F. Fages & M. Hemery (50-50%)

10.2.3. Juries

- Grégory Batt participated in the jurys of
  - PhD Antoine Barizien, Ecole Potytechnique and Institut Pasteur, *Rapporteur*, Paris, May 2019
  - PhD Mathilde Koch, Ecole Polytechnique and INRA, *Examinateur*, Paris, November 2019
- François Fages participated in the jurys of
  - HDR Sabine Pérès, Université Paris-Saclay, *Reviewer*, 18 Nov. 2019
  - HDR David Safranek, Mazaryk Univ., Czech Republic *Reviewer*, July 2019
  - HDR Sriram Krishnamachary, Indraprastha Institute of Information Technology, Delhi, India, *Reviewer*, April 2019
  - PhD Adrien Husson, Université Paris-Diderot, *Rapporteur*, 16 Dec. 2019
  - PhD Jorgelindo da Viega Moreira, Université Paris-Saclay, *Rapporteur*, 18 April 2019

- Master Thesis, Orianne Bargain, Graph matching, theory and SAT implementation, Technische Universität Dresden, Germany, *Superviser* October 2019
- Master Thesis, Zi-Jun Lin, National Taiwan University, *Reviewer*, July 2019
- Master Thesis, Wei-Chih Huang, National Taiwan University, *Reviewer*, July 2019
- Master Thesis, Elisabeth Degrand, Evolving Chemical Reaction Networks, Kungliga tekniska ho¨gskolan, Stockholm, Sweden, *Superviser* April 2019,

10.3. Popularization

10.3.1. Articles and contents

- For a Festschrift in honor of Catuscia Palamidessi, François Fages was unexpectedly inspired to compose a small music score and write a short essay on information leakage in music scores [10].
- Jakob Ruess gave an interview for the Swiss Institute of Bioinformatics (SIB) as a past laureate of the SIB Award.

10.3.2. Interventions

- Eléonore Bellot, Mathieu Hemery and Elise Weill Duflos participated at Fête de la Science at Inria SIF, Oct. 2019
- François Fages has received college school students for a visit to our research team with the question “Can we program any function ?”, answered negatively by proving Cantor’s theorem.
- Mathieu Hemery has animated debates with college school students on Science and Ethics at Inria and Association Arbre des Connaissance, Oct. 2019.

10.3.3. Creation of media or tools for science outreach

François Fages has created several BIOCHAM interactive notebooks that are integrated in the current release of BIOCAHM-4 to illustrate and exercise the main concepts of analog chemical computer and chemical programming, as teached in his Master course at MPRI.

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] S. A. CEPEDA-HUMEREZ, J. RUESS, G. TKAČIK.*Estimating information in time-varying signals*, in "PLoS Computational Biology", September 2019, vol. 15, n^{O} 9, e1007290 [*DOI :* 10.1371/JOURNAL.PCBI.1007290], https://hal.inria.fr/hal-02304413

[2] M. HEMERY, P. FRANÇOIS.*Convergent evolution in silico of biochemical log-response*, in "Journal of Chemical Physics", February 2019, https://hal.inria.fr/hal-02389284

[3] J. RUESS, M. PLEŠKA, C. C. GUET, G. TKAČIK.*Molecular noise of innate immunity shapes bacteria-phage ecologies*, in "PLoS Computational Biology", July 2019 [*DOI :* 10.1371/JOURNAL.PCBI.1007168], https://hal.archives-ouvertes.fr/hal-02229803

International Conferences with Proceedings

[4] E. DEGRAND, M. HEMERY, F. FAGES.*On Chemical Reaction Network Design by a Nested Evolution Algorithm*, in "CMSB 2019 - 17th International Conference on Computational Methods in Systems Biology", Trieste, Italy, LNCS, Springer-Verlag, September 2019, n^{O} 11773, https://hal.inria.fr/hal-02173682

[5] W.-C. HUANG, J.-H. JIANG, F. FAGES, F. MOLINA.*Biochemical Threshold Function Implementation with Zero-Order Ultrasensitivity*, in "BioCAS 2019 - IEEE Biomedical Circuits and Systems Conference", Nara, Japan, IEEE, October 2019, p. 1-4 [*DOI :* 10.1109/BIOCAS.2019.8919176], https://hal.inria.fr/hal-02425761

[6] M. KRYUKOV, A. CARCANO, G. BATT, J. RUESS.*Can optimal experimental design serve as a tool to characterize highly non-linear synthetic circuits?*, in "ECC 2019 - European Control Conference", Naples, Italy, June 2019, https://hal.inria.fr/hal-02304425

[7] J. MARTINELLI, J. GRIGNARD, S. SOLIMAN, F. FAGES.*A Statistical Unsupervised Learning Algorithm for Inferring Reaction Networks from Time Series Data*, in "ICML 2019 - Workshop on Computational Biology", Long Beach, CA, United States, June 2019, https://hal.inria.fr/hal-02163862

[8] J. MARTINELLI, J. GRIGNARD, S. SOLIMAN, F. FAGES.*On Inferring Reactions from Data Time Series by a Statistical Learning Greedy Heuristics*, in "CMSB 2019 - 17th Computational Methods in Systems Biology", Trieste, Italy, L. BORTOLUSSI, G. SANGUINETTI (editors), LNCS, Springer-Verlag, September 2019, n^{o} 11773, https://hal.inria.fr/hal-02173721

[9] E. WEILL, V. ANDREANI, C. ADITYA, P. MARTINON, J. RUESS, G. BATT, F. BONNANS.*Optimal control of an artificial microbial differentiation system for protein bioproductioń*, in "ECC 2019 - European Control Conference", Naples, Italy, June 2019, https://hal.inria.fr/hal-02429963

Scientific Popularization

[10] F. FAGES.*Information Leakage in a Music Score*, in "The Art of Modelling Computational Systems - A Journey from Logic and Concurrency to Security and Privacy - Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday", Lecture Notes in Computer Science, Springer-Verlag, October 2019, vol. Festschrift - LNCS, n^{o} 11760 [*DOI :* 10.1007/978-3-030-31175-9], https://hal.inria.fr/hal-02365478

Other Publications

[11] O. BARGAIN.*Graph matching, theory and SAT implementation*, Technische Universität Dresden, October 2019, https://hal.inria.fr/hal-02339907

[12] E. DEGRAND.*Evolving Chemical Reaction Networks*, Master's Thesis, Kungliga tekniska högskolan (Stockholm), April 2019, p. 1-70, https://hal.inria.fr/hal-02333691

Project-Team M3DISIM

Mathematical and Mechanical Modeling with Data Interaction in Simulations for Medicine

IN COLLABORATION WITH: Laboratoire de Mécanique des Solides

IN PARTNERSHIP WITH:
Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Modeling and Control for Life Sciences

Table of contents

Project-Team M3DISIM

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 June 01

Keywords:

Computer Science and Digital Science:

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A6.3.4. - Model reduction
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization
A6.4.6. - Optimal control
A6.5.1. - Solid mechanics
A6.5.2. - Fluid mechanics
A6.5.4. - Waves
A9.2. - Machine learning

Other Research Topics and Application Domains:

B1.1.3. - Developmental biology
B1.1.8. - Mathematical biology
B1.1.9. - Biomechanics and anatomy
B2.2.1. - Cardiovascular and respiratory diseases
B2.6.2. - Cardiac imaging
B2.6.3. - Biological Imaging

1. Team, Visitors, External Collaborators

Research Scientists

Dominique Chapelle [Inria, Senior Researcher, HDR]
Philippe Moireau [Team leader, Inria, Senior Researcher, HDR]
Radomir Chabiniok [Inria, Advanced Research Position]
Frédérique Clément [Inria, Senior Researcher, HDR]
Sébastien Imperiale [Inria, Researcher]

Faculty Members

Jean-Marc Allain [Ecole polytechnique, Associate Professor, HDR]
Martin Genet [Ecole polytechnique, Associate Professor]
Patrick Le Tallec [Ecole polytechnique, Professor, HDR]

External Collaborators

Matthieu Caruel [Univ Paris-Val de Marne]

Romain Yvinec [INRA]
Maya de Buhan [CNRS]
Didier Lucor [CNRS]
Fabrice Vallée [Assistance publique/Hôpitaux de Paris]
Florent Wijanto [PhD Student, from Jul 2019 until Sep 2019]

Technical Staff

Jerome Diaz [Inria, Engineer]
François Kimmig [Inria, Engineer, from Dec 2019]
Gautier Bureau [EDF, Engineer, until Feb 2019]

PhD Students

Guillaume Ballif [Inria, PhD Student, from Oct 2019]
Ezgi Berberoglu [ETH Zurich, PhD Student]
Federica Caforio [Inria, PhD Student, until Jan 2019]
Chloe Giraudet [Ecole polytechnique, PhD Student]
Marija Gusseva [Inria, PhD Student]
Jona Joachim [Ecole polytechnique, PhD Student, from Nov 2019]
François Kimmig [Ecole polytechnique, PhD Student, until Nov 2019]
Arthur Le Gall [Assistance publique/Hôpitaux de Paris, PhD Student]
Jessica Manganotti [Inria, PhD Student, from Oct 2019]
Cecile Patte [Inria, PhD Student]
Frédérique Robin [Inria, PhD Student, until Sep 2019]
Nicole Tueni [Ecole polytechnique, PhD Student]
Florent Wijanto [Ecole polytechnique, PhD Student, until Jun 2019]

Visiting Scientist

Francesco Regazzoni [Ecole d'ingénieurs, PhD Student, from Jan 2019 until Mar 2019 and from Dec 2019]

2. Overall Objectives

2.1. Overall Objectives

The research carried out in the M3DISIM team has a rather global methodological perspective oriented towards biomechanics, encompassing mathematical modeling and analysis, inverse problems arising from model-data coupling, and the formulation and analysis of effective and reliable numerical procedures adapted to this overall program. We are also very keen on demonstrating the effectiveness and relevance of these methods in actual applications, usually by proof-of-concept studies carried out within various collaborations.

3. Research Program

3.1. Multi-scale modeling and coupling mechanisms for biomechanical systems, with mathematical and numerical analysis

Over the past decade, we have laid out the foundations of a multi-scale 3D model of the cardiac mechanical contraction responding to electrical activation. Several collaborations have been crucial in this enterprise, see below references. By integrating this formulation with adapted numerical methods, we are now able to represent the whole organ behavior in interaction with the blood during complete heart beats. This subject was our first achievement to combine a deep understanding of the underlying physics and physiology and our constant concern of proposing well-posed mathematical formulations and adequate numerical discretizations. In fact, we have shown that our model satisfies the essential thermo-mechanical laws, and in particular the energy balance, and proposed compatible numerical schemes that – in consequence – can be rigorously analyzed, see [6]. In the same spirit, we have formulated a poromechanical model adapted to the blood perfusion in the heart, hence precisely taking into account the large deformation of the mechanical medium, the fluid inertia and moving domain, and so that the energy balance between fluid and solid is fulfilled from the model construction to its discretization, see [7].

3.2. Inverse problems with actual data – Fundamental formulation, mathematical analysis and applications

A major challenge in the context of biomechanical modeling – and more generally in modeling for life sciences – lies in using the large amount of data available on the system to circumvent the lack of absolute modeling ground truth, since every system considered is in fact patient-specific, with possibly non-standard conditions associated with a disease. We have already developed original strategies for solving this particular type of inverse problems by adopting the observer stand-point. The idea we proposed consists in incorporating to the classical discretization of the mechanical system an estimator filter that can use the data to improve the quality of the global approximation, and concurrently identify some uncertain parameters possibly related to a diseased state of the patient. Therefore, our strategy leads to a coupled model-data system solved similarly to a usual PDE-based model, with a computational cost directly comparable to classical Galerkin approximations. We have already worked on the formulation, the mathematical and numerical analysis of the resulting system – see [5] – and the demonstration of the capabilities of this approach in the context of identification of constitutive parameters for a heart model with real data, including medical imaging, see [3].

4. Highlights of the Year

4.1. Highlights of the Year

The team obtained 3 ANR fundings this year: LungManyScale, ODISSE and SIMR.

5. New Software and Platforms

5.1. HeartLab

KEYWORDS: Computational geometry - Image analysis - Cardiac - Health - Simulation

FUNCTIONAL DESCRIPTION: The heartLab software is a library designed to perform both simulation and estimation of the heart mechanical behavior (based on various types of measurements, e.g. images).

Also included are geometric data and tools in the code to define cardiac anatomical models compatible with the simulation requirements in terms of mesh quality, fiber direction data defined within each element, and the referencing necessary for handling boundary conditions and estimation, in particular. These geometries are analytical or come from computerized tomography (CT) or magnetic resonance (MR) image data of humans or animals.

- Participants: Radomir Chabiniok, Gautier Bureau, Martin Genet, Federica Caforio, Ustim Khristenko, Dominique Chapelle and Philippe Moireau
- Contact: Philippe Moireau
- URL: https://raweb.inria.fr/rapportsactivite/RA2013/m3disim/uid14.html

5.2. Verdandi

KEYWORDS: HPC - Model - Software Components - Partial differential equation

FUNCTIONAL DESCRIPTION: Verdandi is a free and open-source (LGPL) library for data assimilation. It includes various such methods for coupling one or several numerical models and observational data. Mainly targeted at large systems arising from the discretization of partial differential equations, the library is devised as generic, which allows for applications in a wide range of problems (biology and medicine, environment, image processing, etc.). Verdandi also includes tools to ease the application of data assimilation, in particular in the management of observations or for a priori uncertainty quantification. Implemented in C++, the library may be used with models implemented in Fortran, C, C++ or Python.

- Participants: Dominique Chapelle, Gautier Bureau, Nicolas Claude, Philippe Moireau and Vivien Mallet
- Contact: Vivien Mallet
- URL: http://verdandi.gforge.inria.fr/

5.3. CardiacLab

KEYWORDS: Cardiovascular and respiratory systems - Matlab - Real time

FUNCTIONAL DESCRIPTION: CardiacLab is a MATLAB toolbox allowing to perform "real-time" cardiac simulations using 0D models of the cardiovascular systems. Its modular development includes (1) a module integrating the mechanical dynamics of the cavity taking into account its particular geometry, (2) a module allowing to choose a micro-model of the cardiac contraction, (3) a module of phase management, (4) a circulation module based on Windkessel models or more advanced 1D flows models, and (5) a perfusion module. The objective of this code is threefold: (1) demonstrate to students, engineers, medical doctors, the interest of modeling in cardiac applications, (2) unify our original modeling developments with the possibility to evaluate them with previous team developments before integrating them into 3D complex formulations, and (3) explore some avenues pertaining to real-time simulat

- Participants: Sebastien Impériale, Martin Genet, Federica Caforio, Ustim Khristenko, Peter Baumgartner, Radomir Chabiniok, François Kimmig and Arthur Le Gall
- Contact: Philippe Moireau
- URL: https://gitlab.inria.fr/M3DISIM/CardiacLab

5.4. MoReFEM

Modeling Research with the Finite Element Method

KEYWORDS: HPC - Multiphysics modelling - Data assimilation

FUNCTIONAL DESCRIPTION: MoReFEM is a HPC finite element library for simulating multiphysics evolution problems like the ones encounter in cardiac modeling (electrophysiology, structure and fluid mechanics, transport-diffusion, wave equations)

- Participants: Philippe Moireau, Patrick Le Tallec, Antoine Olivier, Dominique Chapelle, Ustim Khristenko, François Kimmig, Gautier Bureau and Sébastien Gilles
- Contact: Sébastien Gilles
- URL: https://gitlab.inria.fr/MoReFEM

6. New Results

6.1. Mathematical and Mechanical Modeling

6.1.1. Stochastic modeling of chemical-mechanical coupling in striated muscles

Participants: Matthieu Caruel, Philippe Moireau, Dominique Chapelle [correspondant].

In [18] we propose a chemical–mechanical model of myosin heads in sarcomeres, within the classical description of rigid sliding filaments. In our case, myosin heads have two mechanical degrees-of-freedom (dofs)—one of which associated with the so-called power stroke—and two possible chemical states, i.e., bound to an actin site or not. Our major motivations are twofold: (1) to derive a multiscale coupled chemical–mechanical model and (2) to thus account—at the macroscopic scale—for mechanical phenomena that are out of reach for classical muscle models. This model is first written in the form of Langevin stochastic equations, and we are then able to obtain the corresponding Fokker–Planck partial differential equations governing the probability density functions associated with the mechanical dofs and chemical states. This second form is important, as it allows to monitor muscle energetics and also to compare our model with classical ones, such as the Huxley'57 model to which our equations are shown to reduce under two different types of simplifying assumptions. This provides insight and gives a Langevin form for Huxley'57. We then show how we can calibrate our model based on experimental data—taken here for skeletal muscles—and numerical simulations demonstrate the adequacy of the model to represent complex physiological phenomena, in particular the fast isometric transients in which the power stroke is known to have a crucial role, thus circumventing a limitation of many classical models.

6.1.2. Upscaling of elastic network models

Participant: Patrick Le Tallec.

This work is done in collaboration with Julie Diani from École Polytechnique. The purpose of the approach is to develop general upscaling strategy for deriving macroscopic constitutive laws for rubberlike materials from the knowledge of the network distribution and a mechanical description of the individual chains and of their free energy. It is based on a variational approach in which the microscopic configuration is described by the position of the crosslinks and is obtained not by an affine assumption but by minimizing the corresponding free energy on stochastic large Representative Volume Elements with adequate boundary conditions. This general framework is then approximated by using a microsphere (directional) description of the network and by performing a local minimisation of the network free energy on this simplified configuration space under a maximal advance path kinematic constraint. This approximation framework takes into account anisotropic damage and is extended to handle situations with tube like constraints and stress induced cristallisation. For more detail see [23].

6.1.3. Stochastic construction of surrogate multiphase materials

Participant: Patrick Le Tallec.

Random microstructures of heterogeneous materials play a crucial role in the material macroscopic behavior and in predictions of its effective properties. A common approach to modeling random multiphase materials is to develop so-called surrogate models approximating statistical features of the material. However, the surrogate models used in fatigue analysis usually employ simple microstructure, consisting of ideal geometries such as ellipsoidal inclusions, which generally does not capture complex geometries. In our work , we introduce a simple but flexible surrogate microstructure model for two-phase materials through a level-cut of a Gaussian random field with covariance of Matern class. In addition to the traditional morphology descriptors such as porosity, size and aspect ratio of inclusions, our approach provides control of the regularity of the inclusions interface and sphericity. These parameters are estimated from a small number of real material images using Bayesian inversion. An efficient process of evaluating the samples, based on the Fast Fourier Transform, makes possible the use of Monte-Carlo methods to estimate statistical properties for the quantities of interest in a given material class. This work in progress is done in collaboration with Andrei Constantinescu (École Polytechnique), Ustim Khristenko and Barbara Wohlmuth (Technical University Munich) and Tinsley Oden (University of Texas at Austin). This work has been submitted for publication in an international journal.

6.1.4. Apprehending the effects of mechanical deformations in cardiac electrophysiology – A homogenization approach

Participants: Annabelle Collin [MONC], Sébastien Imperiale, Philippe Moireau, Jean-Frédéric Gerbeau [Inria Siège], Dominique Chapelle [correspondant].

In this work [22], we follow a formal homogenization approach to investigate the effects of mechanical deformations in electrophysiology models relying on a bidomain description of ionic motion at the microscopic level. To that purpose, we extend these microscopic equations to take into account the mechanical deformations, and proceed by recasting the problem in the frame- work of classical two-scale homogenization in periodic media, and identifying the equations satisfied by the first coefficients in the formal expansions. The homogenized equations reveal some interesting effects related to the microstructure – and associated with a specific cell problem to be solved to obtain the macroscopic conductivity tensors – in which mechanical deformations play a non-trivial role, i.e. do not simply lead to a standard bidomain problem posed in the deformed configuration. We then present detailed numerical illustrations of the homogenized model with coupled cardiac electrical-mechanical simulations – all the way to ECG simulations – albeit without taking into account the abundantly-investigated effect of mechanical deformations in ionic models, in order to focus here on other effects. And in fact our numerical results indicate that these other effects are numerically of a comparable order, and therefore cannot be disregarded.

6.1.5. Patient-specific pulmonary mechanics - Modelling and estimation. Application to pulmonary fibrosis.

Participants: Cecile Patte [correspondant], Martin Genet, Dominique Chapelle.

Interstitial pulmonary diseases, like Idiopathic Pulmonary Fibrosis (IPF), affect the alveolar structure of lung tissue, which impacts lung mechanical properties and pulmonary functions. In this work [43], we aim to better understand the pulmonary mechanics in order to improve IPF diagnosis. We developped a poromechanical model for the lung at the organ scale and at the breathing scale. This model is then used to estimate regional mechanical parameters based on clinical data. In the future, this process can be used as an augmented diagnosis tool for clinicians. This work has been presented at the CSMA conference.

6.1.6. Energy preserving cardiac circulation models: formulation, reduction, coupling, inversion, and discretization

Participants: Jessica Manganotti, Philippe Moireau, Sébastien Imperiale [correspondant], Miguel Fernandez [Inria Paris, COMMEDIA].

The modeling of the heart cannot be satisfying if not coupled to the body circulation, and at least to the arterial circulation, which is its direct output “boundary condition”. But more importantly in the clinical context, it is still difficult – ant very invasive – to access the ventricular pressure, which is absolutely necessary for specifying the heart activity. By comparison, more and more devices allow to register non-invasively a distal pressure, for instance at the wrist or the finger, which could be used to estimate the ventricular pressure by inversion of a well adapted arterial circulation model. Such relation is of major interest for clinicians, for example anesthetists, since it could allow real-time monitoring and prediction of the effects of injected drugs during a clinical intervention. Models of the arterial circulation is a well-known subject where dimension reduction has been widely studied for more than half a century. However, the question remains of formulating energy-consistent formulations that can be consistently maintained during the reduction, when coupled to a heart model, and also when discretized. Yet the question is crucial for a better understanding of the physical phenomena of blood flow ejection from the heart as well as the propagation in the arterial network. Moreover, as these models are non-linear, energy-preserving approaches are one of the few tools at our disposal to mathematically justify modeling, discretization or inversion approaches. Finally, inverting this unsteady model for estimation purposes of medical data also benefits from energy-preserving formulation as the inverse approach should also satisfy some stability properties. The subject here is twofold and part of the thesis of J. Manganotti. First we plan to develop accurate models, coupling strategies and robust numerical methods of the arterial network propagation coupled to the heart. Second, we want to develop observer-based strategies that will allow to easily feed these models with measurements in order to perform state estimation of hidden variables or identify key biophysical parameters.

6.1.7. Hierarchical modeling of force generation in cardiac muscle

Participants: Matthieu Caruel, François Kimmig [correspondant].

Performing physiologically relevant simulations of the beating heart in clinical context requires to develop detailed models of the microscale force generation process. These models however may be difficult to implement in practice due to their high computational costs and complex calibration. We propose a hierarchy of three interconnected cardiac muscle contraction models – from the more refined to the more simplified – that are rigorously and systematically related with each other, offering a way to select, for a specific application, the model that yields the best trade-off between physiological fidelity, computational cost and calibration complexity. Our starting model takes into account the stochastic dynamics of the molecular motors force producing conformational changes– and in particular the power stroke – and captures all the timescales of appearing in classical experimental isotonic responses of a heart papillary muscle submitted to rapid load changes. Adiabatic elimination of fast relaxing variables of the stochastic model yields a formulation based on partial differential equations (PDEs) that falls into the family of the Huxley'57 model, while embedding some properties of the process occurring at the fastest timescales. The third family of models is deduced from the PDE model by making minimal assumptions on the parameters, which leads to a computationally light formulation based on ordinary differential equations only. The three models families are compared to the same set of experimental data to systematically assess what physiological indicators can be reproduced or not and how these indicators constrain the model parameters. Finally, we discuss the applicability of these models for heart simulation. This work has been submitted for publication in an international journal.

6.1.8. A relaxed growth modeling framework for controlling growth-induced residual stresses

Participant: Martin Genet.

Background Constitutive models of the mechanical response of soft tissues have been established and are widely accepted, but models of soft tissues remodeling are more controversial. Specifically for growth, one important question arises pertaining to residual stresses: existing growth models inevitably introduce residual stresses, but it is not entirely clear if this is physiological or merely an artifact of the modeling framework. As a consequence, in simulating growth, some authors have chosen to keep growth-induced residual stresses, and others have chosen to remove them. **Methods** In this work, we introduce a novel "relaxed growth" framework allowing for a fine control of the amount of residual stresses generated during tissue growth. It is a direct extension of the classical framework of the multiplicative decomposition of the transformation gradient, to which an additional sub-transformation is introduced in order to let the original unloaded configuration evolve, hence relieving some residual stresses. We provide multiple illustrations of the framework mechanical response, on time-driven constrained growth as well as the strain-driven growth problem of the artery under internal pressure, including the opening angle experiment. **Findings** The novel relaxed growth modeling framework introduced in this paper allows for a better control of growth-induced residual stresses compared to standard growth models based on the multiplicative decomposition of the transformation gradient. **Interpretation** Growth-induced residual stresses should be better handled in soft tissues biomechanical models, especially in patient-specific models of diseased organs that are aimed at augmented diagnosis and treatment optimization. See [27] for more detail.

6.1.9. Multiscale population dynamics in reproductive biology: singular perturbation reduction in deterministic and stochastic models

Participants: Frédérique Clément [correspondant], Romain Yvinec.

During the supervision of a CEMRACS2018 project performed by Céline Bonnet (CMAP) and Keltoum Chahour (LERMA and JLAD), we have described (with Marie Postel, Sorbonne Université and Romain Yvinec, INRA) different modeling approaches for ovarian follicle population dynamics, based on either ordinary (ODE), partial (PDE) or stochastic (SDE) differential equations, and accounting for interactions between follicles [50]. We have put a special focus on representing the population-level feedback exerted by growing ovarian follicles onto the activation of quiescent follicles. We have taken advantage of the timescale difference existing between the growth and activation processes to apply model reduction techniques in the framework of singular perturbations. We have first studied the linear versions of the models to derive theoretical results on the convergence to the limit models. In the nonlinear cases, we have provided detailed numerical evidence of convergence to the limit behavior. We have reproduced the main semi-quantitative features characterizing the ovarian follicle pool, namely a bimodal distribution of the whole population, and a slope break in the decay of the quiescent pool with aging.

6.1.10. Stochastic nonlinear model for somatic cell population dynamics during ovarian follicle activation

Participants: Frédérique Clément [correspondant], Frédérique Robin, Romain Yvinec.

In mammals, female germ cells are sheltered within somatic structures called ovarian follicles, which remain in a quiescent state until they get activated, all along reproductive life. We have investigated the sequence of somatic cell events occurring just after follicle activation [54]. We have introduced a nonlinear stochastic model accounting for the joint dynamics of two cell types, either precursor or proliferative cells. The initial precursor cell population transitions progressively to a proliferative cell population, by both spontaneous and self-amplified processes. In the meantime, the proliferative cell population may start either a linear or exponential growing phase. A key issue is to determine whether cell proliferation is concomitant or posterior to cell transition, and to assess both the time needed for all precursor cells to complete transition and the corresponding increase in the cell number with respect to the initial cell number. Using the probabilistic theory of first passage times, we have designed a numerical scheme based on a rigorous Finite State Projection and coupling techniques to assess the mean extinction time and the cell number at extinction time. We have also obtained analytical formulas for an approximating branching process. We have calibrated the model parameters using an exact likelihood approach using both experimental and in-silico datasets. We have carried out a comprehensive comparison between the initial model and a series of submodels, which help to select the critical cell events taking place during activation. We have finally interpreted these results from a biological viewpoint.

6.1.11. A multiscale mathematical model of cell dynamics during neurogenesis in the mouse cerebral cortex

Participant: Frédérique Clément.

This work is a collaboration with Marie Postel and Sylvie Schneider-Maunoury (Sorbonne Université), Alice Karam (Sorbonne Universités), Guillaume Pézeron (MNHN).
Neurogenesis in the murine cerebral cortex involves the coordinated divisions of two main types of progenitor cells, whose numbers, division modes and cell cycle durations set up the final neuronal output. In this work [33] we aim at understanding the respective roles of these factors in the neurogenesis process, we have combined experimental in vivo studies with mathematical modeling and numerical simulations of the dynamics of neural progenitor cells. A special focus is put on the population of intermediate progenitors (IPs), a transit amplifying progenitor type critically involved in the size of the final neuron pool. A multiscale formalism describing IP dynamics allows one to track the progression of cells along the subsequent phases of the cell cycle, as well as the temporal evolution of the different cell numbers. Our model takes into account the dividing apical progenitors (AP) engaged into neurogenesis, both neurogenic and proliferative IPs, and the newborn neurons. The transfer rates from one population to another are subject to the mode of division (symmetric, asymmetric, neurogenic) and may be time-varying. The model outputs have been successfully fitted to experimental cell numbers from mouse embryos at different stages of cortical development, taking into account IPs and neurons, in order to adjust the numerical parameters. Applying the model to a mouse mutant for Ftm/Rpgrip1l, a gene involved in human ciliopathies with severe brain abnormalities, reveals a shortening of the neurogenic period associated with an increased influx of newborn IPs from apical progenitors at mid-neurogenesis. Additional information is provided on cell kinetics, such as the mitotic and S phase indexes, and neurogenic fraction. Our model can be used to study other mouse mutants with cortical neurogenesis defects and can be adapted to study the importance of progenitor dynamics in cortical evolution and human diseases.

6.2. Numerical Methods

6.2.1. Numerical analysis for an energy-stable total discretization of a poromechanics model with inf-sup stability

Participants: Dominique Chapelle [correspondant], Philippe Moireau.

In this joint work with Bruno Burtschell [16], we consider a previously proposed general nonlinear poromechanical formulation, and we derive a linearized version of this model. For this linearized model, we obtain an existence result and we propose a complete discretization strategy–in time and space–with a special concern for issues associated with incompressible or nearly-incompressible behavior. We provide a detailed mathematical analysis of this strategy, the main result being an error estimate uniform with respect to the compressibility parameter. We then illustrate our approach with detailed simulation results and we numerically investigate the importance of the assumptions made in the analysis, including the fulfillment of specific inf-sup conditions.

6.2.2. Conservative and entropy controlled remap for multi-material ALE simulations

Participant: Patrick Le Tallec.

For many multi-material problems such as fluid-structure interaction, impact or implosion problems, materials are in very large strains due to their nature or to the applied forces. In our situations of interest, we also have a strong coupling between energy and momentum conservation laws, due to intense transfers between internal and kinetic energies and to strong advection effects. Such situations are classically governed by the Euler's equations, written in Lagrangian form, and using a multi-material, single velocity framework, but their numerical solution demands a strict control of energy conservation and entropy production, which is hard to achieve in situations where dynamic remeshing is mandatory. In this framework, our approach deals with the analysis of the impact of a second-order staggered remap using an intersection-based approach on conservation properties and on the entropy control. We show that an accurate remap with exact mesh intersections and exact integrations affects both the momentum and the kinetic energy because of node mass re-localizations and node velocity remap. We propose therefore a staggered remapping strategy in order to take into account these discrepancies at a low computational cost. While preserving the strict conservation of total energy, our strategy allows to recover a proper entropy control at the expense of strict momentum conservation and monotonicity losses. This work [32] is done in collaboration with Alexandra Claisse (CEA DAM) and Alexis Marboeuf (École Polytechnique and CEA DAM).

6.2.3. Multipatch isogeometric analysis for complex structures

Participant: Patrick Le Tallec.

This work – done in collaboration with Nicolas Adam (École Polytechnique and PSA) and Malek Zarroug (PSA) – introduces, analyzes and validates isogeometric mortar methods for the solution of thick shells problems which are set on a multipatch geometry. It concerns industrial parts of complex geometries for which the effects of transverse shear cannot be neglected. For this purpose, Reissner-Mindlin model was retained and rotational degrees of freedom (DOF) of the normal are taken into account. A particular attention is devoted to the introduction of a proper formulation of the coupling conditions at patches interfaces, with a particular interest on augmented lagrangian formulations, to the choice and validation of mortar spaces, and to the derivation of adequate integration rules. The relevance of the proposed approach is assessed numerically on various significative examples of industrial relevance. This work has been submitted for publication in an international journal.

6.2.4. Mathematical and numerical study of transient wave scattering by obstacles with the Arlequin Method

Participant: Sébastien Imperiale.

In this work [14] we extend the Arlequin method to overlapping domain decomposition technique for transient wave equation scattering by obstacles. The main contribution of this work is to construct and analyze from the continuous level up to the fully discrete level some variants of the Arlequin method. The constructed discretizations allow to solve wave propagation problems while using non-conforming and overlapping meshes for the background propagating medium and the surrounding of the obstacle respectively. Hence we obtain a flexible and stable method in terms of the space discretization – an inf-sup condition is proven – while the stability of the time discretization is ensured by energy identities.

6.2.5. Construction and analysis of fourth-order, energy consistent, family of explicit time discretizations for dissipative linear wave equations

Participants: Juliette Chabassier [MAGIQUE-3D], Julien Diaz [MAGIQUE-3D], Sébastien Imperiale [correspondant].

This work and the corresponding article [19], deal with the construction of a family of fourth order, energy consistent, explicit time discretizations for dissipative linear wave equations. The schemes are obtained by replacing the inversion of a matrix, that comes naturally after using the technique of the Modified Equation on the second order Leap Frog scheme applied to dissipative linear wave equations, by explicit approximations of its inverse. The stability of the schemes are studied using an energy analysis and a convergence analysis is carried out. Numerical results in 1D illustrate the space/time convergence properties of the schemes and their efficiency is compared to more classical time discretizations.

6.2.6. Energy decay and stability of a perfectly matched layer For the wave equation

Participants: Sébastien Imperiale [correspondant], Maryna Kachanovska [POEMS].

We follow a previous work where PML formulations was proposed for the wave equation in its standard second-order form. In the present work [15], energy decay and L^2 stability bounds in two and three space dimensions are rigorously proved both for continuous and discrete formulations with constant damping coefficients. Numerical results validate the theory.

6.2.7. A high-order spectral element fast Fourier transform for the poisson equation

Participants: Federica Caforio, Sébastien Imperiale [correspondant].

The aim of this work [17] is to propose a novel, fast solver for the Poisson problem discretised with High-Order Spectral Element Methods (HO-SEM) in canonical geometries (rectangle in 2D, rectangular parallelepiped in 3D). This method is based on the use of the Discrete Fourier Transform to reduce the problem to the inversion of the symbol of the operator in the frequency space. The proposed solver is endowed with several properties. First, it preserves the efficiency of the standard FFT algorithm; then, the matrix storage is drastically reduced (in particular, it is independent of the space dimension); a pseudo-explicit Singular Value Decomposition (SVD) is used for the inversion of the symbols; finally, it can be extended to non-periodic boundary conditions. Furthermore, due to the underlying HO-SEM discretisation, the multi-dimensional symbol of the operator can be efficiently computed from the one-dimensional symbol by tensorisation.

6.2.8. Thermodynamic properties of muscle contraction models and associated discrete-time principles

Participants: François Kimmig, Dominique Chapelle [correspondant], Philippe Moireau.

Considering a large class of muscle contraction models accounting for actin-myosin interaction, we present a mathematical setting in which solution properties can be established, including fundamental thermodynamic balances. Moreover, we propose a complete discretization strategy for which we are also able to obtain discrete versions of the thermodynamic balances and other properties. Our major objective is to show how the thermodynamics of such models can be tracked after discretization, including when they are coupled to a macroscopic muscle formulation in the realm of continuum mechanics. Our approach allows to carefully identify the sources of energy and entropy in the system, and to follow them up to the numerical applications. See [30] for more detail.

6.2.9. Mechanical and imaging models-based image registration

Participants: Radomir Chabiniok, Martin Genet [correspondant].

Image registration plays an increasingly important role in many fields such as biomedical or mechanical engineering. Generally speaking, it consists in deforming a (moving) source image to match a (fixed) template image. Many approaches have been proposed over the years; if new model-free machine learning-based approaches are now beginning to provide robust and accurate results, extracting motion from images is still most commonly based on combining some statistical analysis of the images intensity and some model of the underlying deformation as initial guess or regularizer. These approaches may be efficient even for complex type of motion; however, any artifact in the source image (e.g., partial voluming, local decrease of signal-to-noise ratio or even local signal void), drastically deteriorates the registration. This work introduces a novel approach of extracting motion from biomedical image series, based on a model of the imaging modality. It is, to a large extent, independent of the type of model and image data – the pre-requisite is to incorporate biomechanical constraints into the motion of the object (organ) of interest and being able to generate data corresponding to the real image, i.e., having an imaging model at hand. We will illustrate the method with examples of synthetically generated 2D tagged magnetic resonance images. This work was presented at the VipIMAGE 2019 conference. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD. See [44] for more detail. This work was done in collaboration with Katerina Skardova (Czech Technical University in Prague) and Matthias Rambausek (École Polytechnique).

6.2.10. Validation of finite element image registration-based cardiac strain estimation from magnetic resonance images

Participants: Martin Genet [correspondant], Philippe Moireau.

Accurate assessment of regional and global function of the heart is an important readout for the diagnosis and routine evaluation of cardiac patients. Indeed, recent clinical and experimental studies suggest that compared to global metrics, regional measures of function could allow for more accurate diagnosis and early intervention for many cardiac diseases. Although global strain measures derived from tagged magnetic resonance (MR) imaging have been shown to be reproducible for the majority of image registration techniques, the measurement of regional heterogeneity of strain is less robust. Moreover, radial strain is underestimated with the current techniques even globally. Finite element (FE)-based techniques offer a mechanistic approach for the regularization of the ill-posed registration problem. This work presents the validation of a recently proposed FE-based image registration method with mechanical regularization named equilibrated warping. For this purpose, synthetic 3D-tagged MR images are generated from a reference biomechanical model of the left ventricle (LV). The performance of the registration algorithm is consequently tested on the images with different signal-to-noise ratios (SNRs), revealing the robustness of the method. See [35] for more detail.

6.3. Inverse Problems

6.3.1. Analysis of an observer strategy for initial state reconstruction of wave-like systems in unbounded domains

Participants: Sébastien Imperiale, Philippe Moireau [correspondant].

In [29] we are interested in reconstructing the initial condition of a wave equation in an unbounded domain configuration from measurements available in time on a subdomain. To solve this problem, we adopt an iterative strategy of reconstruction based on observers and time reversal adjoint formulations. We prove the convergence of our reconstruction algorithm with perfect measurements and its robustness to noise. Moreover, we develop a complete strategy to practically solve this problem on a bounded domain using artificial transparent boundary conditions to account for the exterior domain. Our work then demonstrates that the consistency error introduced by the use of approximate transparent boundary conditions is compensated by the stabilisation properties obtained from the use of the available measurements, hence allowing to still be able to reconstruct the unknown initial condition.

6.3.2. Analysis and numerical simulation of an inverse problem for a structured cell population dynamics model

Participants: Frédérique Clément, Frédérique Robin [correspondant].

We have studied (with Béatrice Laroche, INRA) a multiscale inverse problem associated with a multi-type model for age structured cell populations [20] (see also [21] for another application). In the single type case, the model is a McKendrick-VonFoerster like equation with a mitosis-dependent death rate and potential migration at birth. In the multi-type case, the migration term results in a unidirectional motion from one type to the next, so that the boundary condition at age 0 contains an additional extrinsic contribution from the previous type. We consider the inverse problem of retrieving microscopic information (the division rates and migration proportions) from the knowledge of macroscopic information (total number of cells per layer), given the initial condition. We have first shown the well-posedness of the inverse problem in the single type case using a Fredholm integral equation derived from the characteristic curves, and we have used a constructive approach to obtain the lattice division rate, considering either a synchronized or non-synchronized initial condition. We have taken advantage of the unidirectional motion to decompose the whole model into nested submodels corresponding to self-renewal equations with an additional extrinsic contribution. We have again derived a Fredholm integral equation for each submodel and deduced the well-posedness of the multi-type inverse problem. In each situation, we illustrate numerically our theoretical results.

6.3.3. Inverse problem based on data assimilation approaches for protein aggregation

Participants: Philippe Moireau [correspondant], Cécile Della Valle [MAMBA], Marie Doumic [MAMBA].

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation. In a depolarization configuration, we here extend some previous results obtained during the PhD Thesis of A. Armiento. Now, the depolarization rate is time-dependent or in the presence of an additional vanishing viscosity term. We continue to develop our framework mixing inverse problems methodologies and optimal control approaches typically encountered in data assimilation, allowing to justify mathematically the methods but also to adopt efficient numerical strategies. Publications of this work will be soon submitted.

6.3.4. Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment

Participants: Annabelle Collin [MONC], Philippe Moireau [correspondant].

Data-driven wildfire spread modeling is emerging as a cornerstone for forecasting real-time fire behavior using thermal-infrared imaging data. One key challenge in data assimilation lies in the design of an adequate measure to represent the discrepancies between observed and simulated firelines (or "fronts"). A first approach consists in adopting a Lagrangian description of the flame front and in computing a Euclidean distance between simulated and observed fronts by pairing each observed marker with its closest neighbor along the simulated front. However, this front marker registration approach is difficult to generalize to complex front topology that can occur when fire propagation conditions are highly heterogeneous due to topography, biomass fuel and micrometeorology. To overcome this issue, we investigate in this paper an object-oriented approach derived from the Chan–Vese contour fitting functional used in image processing. The burning area is treated as a moving object that can undergo shape deformations and topological changes. We combine this non-Euclidean measure with a state estimation approach (a Luenberger observer) to perform simulations of the time-evolving fire front location driven by discrete observations of the fireline. We apply this object-oriented data assimilation method to the three-hectare RxCADRE S5 field-scale experiment. This collaboration with CERFACS (M. Rochoux) and University of Maryland (C. Zhang and A. Trouvé) led to a publication [34] in the Proceedings of the Combustion Institute.

6.3.5. Model assessment through data assimilation of realistic data in cardiac electrophysiology

Participants: Antoine Gerard [CARMEN], Annabelle Collin [MONC], Gautier Bureau, Philippe Moireau [correspondant], Yves Coudière [CARMEN].

We consider a model-based estimation procedure – namely a data assimilation algorithm – of the atrial depolarization state of a sub- ject using data corresponding to electro-anatomical maps. Our objective is to evaluate the sensitivity of such a model-based reconstruction with respect to model choices. The followed data assimilation approach is capable of using electrical activation times to adapt a monodomain model simulation, thanks to an ingenious model-data fitting term inspired from image processing. The resulting simulation smoothes and completes the activation maps when they are spatially incomplete. Moreover, conductivity parameters can also be inferred. The model sensitivity assessment is performed based on synthetic data generated with a validated realistic atria model and then inverted using simpler modeling ingredients. In particular, the impact of the muscle fibers definition and corresponding anisotropic conductivity parameters is studied. Finally, an application of the method to real data is presented, showing promising results. This collaborative work has been published, see [37].

6.4. Experimental Assessments

6.4.1. Combination of traction assays and multiphoton imaging to quantify skin biomechanics

Participant: Jean-Marc Allain.

An important issue in tissue biomechanics is to decipher the relationship between the mechanical behavior at macroscopic scale and the organization of the collagen fiber network at microscopic scale. We have formalized a definitive protocol [46] to combine traction assays with multiphoton microscopy in ex vivo murine skin. This multiscale approach provides simultaneously the stress/stretch response of a skin biopsy and the collagen reorganization in the dermis by use of second harmonic generation (SHG) signals and appropriate image processing.

6.4.2. Monitoring dynamic collagen reorganization during skin stretching with fast polarization-resolved second harmonic generation imaging

Participant: Jean-Marc Allain.

The mechanical properties of biological tissues are strongly correlated to the specific distribution of their collagen fibers. Monitoring the dynamic reorganization of the collagen network during mechanical stretching is however a technical challenge, because it requires mapping orientation of collagen fibers in a thick and deforming sample. In this work [24], a fast polarization-resolved second harmonic generation microscope is implemented to map collagen orientation during mechanical assays. This system is based on line-to-line switching of polarization using an electro-optical modulator and works in epi-detection geometry. After proper calibration, it successfully highlights the collagen dynamic alignment along the traction direction in ex vivo murine skin dermis. This microstructure reorganization is quantified by the entropy of the collagen orientation distribution as a function of the stretch ratio. It exhibits a linear behavior, whose slope is measured with a good accuracy. This approach can be generalized to probe a variety of dynamic processes in thick tissues.

6.4.3. Multiscale characterisation of skin mechanics through in-situ imaging

Participant: Jean-Marc Allain.

The complex mechanical properties of skin have been studied intensively over the past decades. They are intrinsically linked to the structure of the skin at several length scales, from the macroscopic layers (epidermis, dermis and hypodermis) down to the microstructural organization at the molecular level. Understanding the link between this microscopic organization and the mechanical properties is of significant interest in the cosmetic and medical fields. Nevertheless, it only recently became possible to directly visualize the skin's microstructure during mechanical assays, carried out on the whole tissue or on isolated layers. These recent observations have provided novel information on the role of structural components of the skin in its mechanical properties, mainly the collagen fibers in the dermis, while the contribution of others, such as elastin fibers, remains elusive. We performed in [45] a systematic review of the current methods used to observe skin's microstructure during a mechanical assay, along with their strengths and limitations, as well as a review of the unique information they provide on the link between structure and function of the skin.

6.4.4. Root Hair Sizer: an algorithm for high throughput recovery of different root hair and root developmental parameters

Participant: Jean-Marc Allain.

The root is an important organ for water and nutrient uptake, and soil anchorage. It is equipped with root hairs (RHs) which are elongated structures increasing the exchange surface with the soil. RHs are also studied as a model for plant cellular development, as they represent a single cell with specific and highly regulated polarized elongation. For these reasons, it is useful to be able to accurately quantify RH length employing standardized procedures. Methods commonly employed rely on manual steps and are therefore time consuming and prone to errors, restricting analysis to a short segment of the root tip. Few partially automated methods have been reported to increase measurement efficiency. However, none of the reported methods allow an accurate and standardized definition of the position along the root for RH length measurement, making data comparison difficult. In this work [28] we are developing an image analysis algorithm that semi-automatically detects RHs and measures their length along the whole differentiation zone of roots. This method, implemented as a simple automated script in ImageJ/Fiji software that we termed Root Hair Sizer, slides a rectangular window along a binarized and straightened image of root tips to estimate the maximal RH length in a given measuring interval. This measure is not affected by heavily bent RHs and any bald spots. RH length data along the root are then modelled with a sigmoidal curve, generating several biologically significant parameters such as RH length, positioning of the root differentiation zone and, under certain conditions, RH growth rate. Image analysis with Root Hair Sizer and subsequent sigmoidal modelling of RH length data provide a simple and efficient way to characterize RH growth in different conditions, equally suitable to small and large scale phenotyping experiments.

6.4.5. Calcium and plasma membrane force-gated ion channels behind development

Participant: Jean-Marc Allain.

During development, tissues are submitted to high variation of compression and tension forces. The roles of the cell wall, the cytoskeleton, the turgor pressure and the cell geometry during this process have received due attention. In contrast, apart from its role in the establishment of turgor pressure, the involvement of the plasma membrane as a transducer of mechanical forces during development has been under studied. Force-gated (FG) or Mechanosensitive (MS) ion channels embedded in the bilayer represent 'per se' archetypal mechanosensor able to directly and instantaneously transduce membrane forces into electrical and calcium signals. We reviewed in [26] how their fine-tuning, combined with their ability to detect micro-curvature and local membrane tension, allows FG channels to transduce mechanical cues into developmental signals.

6.5. Clinical Applications

6.5.1. Cardiac displacement tracking with data assimilation combining a biomechanical model and an automatic contour detection

Participants: Radomir Chabiniok, Gautier Bureau, Dominique Chapelle, Philippe Moireau [correspondant].

Data assimilation in computational models represents an essential step in building patient-specific simulations. This work aims at circumventing one major bottleneck in the practical use of data assimilation strategies in cardiac applications, namely, the difficulty of formulating and effectively computing adequate data-fitting term for cardiac imaging such as cine MRI. We here provide a proof-of-concept study of data assimilation based on automatic contour detection. The tissue motion simulated by the data assimilation framework is then assessed with displacements extracted from tagged MRI in six subjects, and the results illustrate the performance of the proposed method, including for circumferential displacements, which are not well extracted from cine MRI alone. This work was presented at the Functional Imaging and Modeling of Heart Conference (FIMH2019, Bordeaux, France) and published in [36].

6.5.2. Minimally-invasive estimation of patient-specific end-systolic elastance using a biomechanical heart model

Participants: Arthur Le Gall, Fabrice Vallée, Dominique Chapelle, Radomir Chabiniok [correspondant].

The end-systolic elastance (E_{es}) – the slope of the end-systolic pressure-volume relationship (ESPVR) at the end of ejection phase – has become a reliable indicator of myocardial functional state. The estimation of E_{es} by the original multiple-beat method is invasive, which limits its routine usage. By contrast, non-invasive single-beat estimation methods, based on the assumption of the linearity of ESPVR and the uniqueness of the normalised time-varying elastance curve $E^N(t)$ across subjects and physiology states, have been applied in a number of clinical studies. It is however known that these two assumptions have a limited validity, as ESPVR can be approximated by a linear function only locally, and $E^N(t)$ obtained from a multi-subject experiment includes a confidence interval around the mean function. Using datasets of 3 patients undergoing general anaesthesia (each containing aortic flow and pressure measurements at baseline and after introducing a vasopressor noradrenaline), we first study the sensitivity of two single-beat methods — by Sensaki et al. and by Chen et al. — to the uncertainty of $E^N(t)$. Then, we propose a minimally-invasive method based on a patient-specific biophysical modelling to estimate the whole time-varying elastance curve $E^{model}(t)$. We compare E_{es}^{model} with the two single-beat estimation methods, and the normalised varying elastance curve $E^{N,model}(t)$ with $E^N(t)$ from published physiological experiments. This work was presented at the Functional Imaging and Modeling of Heart conference (FIMH2019, Bordeaux, France) and published in [38].

6.5.3. Model-based indices of early-stage cardiovascular failure and its therapeutic management in Fontan patients

Participant: Radomir Chabiniok.

Investigating the causes of failure of Fontan circulation in individual patients remains challenging despite detailed combined invasive cardiac catheterisation and magnetic resonance (XMR) exams at rest and during stress. In this work, we use a biomechanical model of the heart and Fontan circulation with the components of systemic and pulmonary beds to augment the diagnostic assessment of the patients undergoing the XMR stress exam. We apply our model in 3 Fontan patients and one biventricular “control" case. In all subjects, we obtained important biophysical factors of cardiovascular physiology – contractility, contractile reserve and changes in systemic and pulmonary vascular resistance – which contribute to explaining the mechanism of failure in individual patients. Finally, we used the patient-specific model of one Fontan patient to investigate the impact of changes in pulmonary vascular resistance, aiming at in silico testing of pulmonary vasodilation treatments. This work (in collaboration with Bram Ruijsink and Kuberan Pushparajah from St Thomas Hospital, King’s College London) was presented at the Functional Imaging and Modeling of Heart conference (FIMH2019, Bordeaux, France) and published in [40]. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.4. Dobutamine stress testing in patients with Fontan circulation augmented by biomechanical modeling

Participants: Philippe Moireau, Dominique Chapelle, Radomir Chabiniok [correspondant].

Understanding (patho)physiological phenomena and mechanisms of failure in patients with Fontan circulation — a surgically established circulation for patients born with a functionally single ventricle — remains challenging due to the complex hemodynamics and high inter-patient variations in anatomy and function. In this work, we present a biomechanical model of the heart and circulation to augment the diagnostic evaluation of Fontan patients with early-stage heart failure. The proposed framework employs a reduced-order model of heart coupled with a simplified circulation including venous return, creating a closed-loop system. We deploy this framework to augment the information from data obtained during combined cardiac catheterization and magnetic resonance exams (XMR), performed at rest and during dobutamine stress in 9 children with Fontan circulation and 2 biventricular controls. We demonstrate that our modeling framework enables patient-specific investigation of myocardial stiffness, contractility at rest, contractile reserve during stress and changes in vascular resistance. Hereby, the model allows to identify key factors underlying the pathophysiological response to stress in these patients. In addition, the rapid personalization of the model to patient data and fast simulation of cardiac cycles makes our framework directly applicable in a clinical workflow. We conclude that the proposed modeling framework is a valuable addition to the current clinical diagnostic XMR exam that helps to explain patient-specific stress hemodynamics and can identify potential mechanisms of failure in patients with Fontan circulation. This work has been submitted for publication in an international journal. This work (in collaboration with Bram Ruijsink and Kuberan Pushparajah from St Thomas Hospital, King's College London and Tarique Hussain, UT Southwestern Medical Center Dallas) also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.5. Signed-distance function based non-rigid registration of image sequences with varying image intensity

Participant: Radomir Chabiniok.

In this work we deal with non-rigid registration of the image series acquired by the Modified Look-Locker Inversion Recovery (MOLLI) magnetic resonance imaging sequence, which is used for a pixel-wise estimation of T_1 relaxation time. The spatial registration of the images within the series is necessary to compensate the patient's imperfect breath-holding. The evolution of intensities and a large variation of the image contrast within the MOLLI image series, together with the myocardium of left ventricle (the object of interest) typically not being the most distinct object in the scene, makes the registration challenging. We propose a method for locally adjusted optical flow-based registration of multimodal images, which uses the segmentation of the object of interest and its representation by the signed-distance function. We describe all the components of the proposed OF^{dist} method and their implementation. The OF^{dist} method is then compared to the performance of a standard mutual information maximization-based registration method, applied either to the original image (MIM) or to the signed-distance function (MIM^{dist}). Several experiments with synthetic and real MOLLI images are carried out. On synthetic image with a single object, MIM performed the best, while OF^{dist} and MIM^{dist} provided better results on synthetic images with more than one object and on real images. When applied to signed-distance function of two objects of interest, MIM^{dist} provided a larger registration error (but more homogeneously distributed) compared to OF^{dist}. For the real MOLLI image sequence with left ventricle pre-segmented using level-set method, the proposed OF^{dist} registration performed the best, as is demonstrated visually and by measuring the increase of mutual information in the object of interest and its neighborhood. This collaborative work (Katerina Skardova, Czech Technical University, Institute of Clinical and Experimental Medicine in Prague) has been submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.6. Estimation of left ventricular pressure-volume loop using hemodynamic monitoring augmented by a patient-specific biomechanical model. An observational study

Participants: Arthur Le Gall, Fabrice Vallée, Dominique Chapelle, Radomir Chabiniok [correspondant].

Background During general anaesthesia, direct analysis of the arterial pressure or aortic flow waveforms may be confusing in complex haemodynamic situations. Patient-specific biomechanical modelling allows to simulate Pressure-Volume (PV) loops and obtain functional indicators of the cardiovascular (CV) system, such as ventricular-arterial coupling (Vva), cardiac efficiency (CE) or myocardial contractility. It therefore augments the information obtained by monitoring and could help in medical decision-making. **Methods** Patients undergoing GA for neuroradiological procedure were included in this prospective observational study. A biomechanical model of heart and vasculature specific to each patient was built using transthoracic echocardiography and aortic pressure and flow signals. If intraoperative hypotension (IOH) appeared, diluted noradrenaline (NOR) was administered and the model readjusted. **Results** The model was calibrated for 29 (64%) normotensive and for 16 (36%) hypotensive patients before and after NOR administration. The simulated mean aortic pressure (MAP) and stroke volume (SV) were equivalent to the measurements (Percentage Error: 6% for MAP and 18% for SV) in all 45 datasets at baseline. After NOR administration, the percentage of concordance with 10% exclusion zone between measurement and simulation was > 95% for both MAP and SV. The modelling results showed a decreased Vva (0.64 ± 0.37 vs 0.88 ± 0.43; p=0.039), and an increased CE (0.8 ± 0.1 vs 0.73 ± 0.11; p=0.042) in hypotensive as compared with normotensive patients. After NOR administration, Vva increased by $92 \pm 101\%$, CE decreased by $13 \pm 11\%$ (p<0.001 for both) and contractility increased by $14 \pm 11\%$ (p=0.002). **Conclusions** The numerical models built for individual patients were applied to estimate patients' PV loops and functional indicators of CV system during haemodynamic alterations and during restoration by NOR. This study demonstrates the feasibility of patient-specific cardiovascular modelling using clinical data readily available during GA and paves the way for model-augmented haemodynamic monitoring at operating theatres and intensive care units. This work is about to be submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.7. *Investigation of phase contrast magnetic resonance imaging underestimation of turbulent flow through the aortic valve phantom: Experimental and computational study by using lattice Boltzmann method*

Participant: Radomir Chabiniok.

Work in collaboration with Radek Fucik, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague.

Objective The accuracy of phase-contrast magnetic resonance imaging (PC-MRI) measurement is investigated using a computational fluid dynamics (CFD) model with the objective to determine the magnitude of the flow underestimation due to turbulence behind a narrowed valve in a phantom experiment. **Materials and Methods** An acrylic stationary flow phantom is used with three insertable plates mimicking aortic valvular stenoses of varying degrees. Positive and negative horizontal fluxes are measured at equidistant slices using standard PC-MRI sequences by 1.5T and 3T systems. The CFD model is based on the 3D lattice Boltzmann method (LBM). The experimental and simulated data are compared using the Bland-Altman-derived limits of agreement. Based on the LBM results, the turbulence is quantified and confronted with the level of flow underestimation. **Results** Matching results of PC-MRI flow were obtained for valves up to moderate stenosis on both field strengths. The flow magnitude through a severely stenotic valve was underestimated due to signal void in the regions of turbulent flow behind the valve, consistently with the level of quantified turbulence intensity. **Discussion** Flow measured by PC-MRI is affected by noise and turbulence. LBM can simulate turbulent flow efficiently and accurately, it has therefore the potential to improve clinical interpretation of PC-MRI. This collaborative work (Czech Technical University, Institute of Clinical and Experimental Medicine in Prague and Inria) has been submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.8. *Left ventricular torsion obtained using equilibrated warping in patients with repaired Tetralogy of Fallot*

Participants: Martin Genet, Radomir Chabiniok [correspondant].

Work in collaboration with Katerina Skardova, Department of Mathematics, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague and Tarique Hussain UT Southwestern Medical Center Dallas.

Background Patients after surgical repair of Tetralogy of Fallot (rTOF) have right ventricular (RV) dysfunction and may subsequently suffer a decrease in left ventricular (LV) function. Previous studies evaluating the assessment of LV torsion have shown poor reproducibility using cardiovascular magnetic resonance imaging (CMR). The aim of our study is to evaluate a novel finite element method of image registration to assess LV torsion in patients with rTOF and explore the relationship between LV torsion and cardiac parameters routinely obtained with CMR. **Methods** The assessment of torsion is based on the finite element method for image registration, and the equilibrium gap principle for problem regularization, known as equilibrated warping developed by M. Genet (Inria Saclay). It has been shown to be able to predict global torsion in regular cine images as well in 3D tagged images, despite low contrast. Seventy-six cases of rTOF and ten controls were included. The group of control patients were assessed for reproducibility using equilibrated warping and standard tissue tracking software (cvi42, version 5.10.1, Calgary, Canada). RV end-systolic volume (RVESV), RV end-diastolic volume (RVEDV), RV ejection fraction (RVEF), LVESV, LVEDV, LVEF, LV peak systolic torsion and peak systolic torsion gradient (normalized by mesh length) were obtained for each patient with rTOF. Patients were dichotomized into two groups: those with normal torsion (systolic basal clockwise rotation and apical counterclockwise rotation, representative example is shown in Image 1) and those with loss of torsion, defined as a reversal of normal systolic basal clockwise rotation (representative example is shown in Image 2). **Results** Torsion by equilibrated warping was successfully obtained in 68 of 76 (89%) patients with rTOF and 9 of 10 (90%) normal controls. For equilibrated warping, the intra- and inter-observer coefficients of variation were 0.095 and 0.117, respectively; compared to 0.668 and 0.418 for tissue tracking by standard clinical software. The intra- and inter-observer intraclass correlation coefficients for equilibrated warping were 0.862 and 0.831, respectively; compared to 0.250 and 0.621 for tissue tracking. Loss of torsion was noted in 32 of the 68 (47%) patients with rTOF and there was a significant difference in peak systolic torsion gradient between patients with normal torsion and loss of torsion. There was no difference in LV or RV volumes or function between these groups. **Conclusion** The equilibrated warping method of image registration to assess LV torsion is feasible in patients with rTOF and shows good reliability. Loss of torsion is common in patients with rTOF. In our study, there was no significant association between loss of torsion and other ventricular parameters indicative of a worsening cardiac condition. Future studies committed to the long- term follow-up of this population are needed to assess the role of torsion in predicting ventricular dysfunction and death. This work was accepted for presentation at SCMR conference 2020 (Society for Cardiovascular Magnetic Resonance). It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.9. Volume administration protocol to assess ventricular mechanics during interventional cardiac magnetic resonance procedures

Participant: Radomir Chabiniok.

Work in collaboration with Joshua Greer and Tarique Hussain UT Southwestern Medical Center Dallas.

Background Failure in Fontan circulation occurs with supposed normal ventricular systolic and diastolic function, including normal ventricular end-diastolic pressures and ventricular ejection fraction. This highlights the difficulty in assessing systolic and diastolic ventricular function in patients with single ventricle physiology. Interventional cardiac magnetic resonance (CMR) provides an opportunity for simultaneous acquisition of pressure and volume measurements that may lend itself well to analysis of ventricular mechanics in this population. We aim to develop a protocol of volume administration to assess ventricular pressure and volume during the cardiac cycle to construct pressure-volume loops under different loading conditions and perform their biomechanical interpretation. **Methods** This is a single center prospective study conducted on single ventricle patients with Glenn or Fontan circulation referred for interventional CMR procedures. With a catheter advanced into the ventricle, a pressure tracing and a cine sequence accelerated by kt-BLAST is obtained. Two 2.5 mL/kg fluid boluses are then rapidly administered into the catheter sheath with repeated acquisition of the pressure tracing and cine imaging immediately following each. Cine images are post-processed after the procedure to obtain ventricular volumes. The data are combined to construct pressure-volume loops and plot the end-diastolic pressure-volume relationship (EDPVR). **Results** The protocol has been performed in six patients. Ventricular end-diastolic pressure readings increased by a median of 2.5 mmHg (range of 1-3 mmHg) after the first volume administration and a median of 1.5 mmHg (range of 1-8 mmHg) after the second volume administration. Ventricular end-diastolic volumes increased by a median of 4.1 mL (range of 1.7-19.3 mL) after the first volume administration and a median of 1.5 mL (range of 0.4-24.2 mL) after the second volume administration. The data obtained during simultaneous volume and pressure measurements allowed for the construction of ventricular pressure-volume loops. Ventricular stroke work increased by a median of 0.0825 Joules (range of 0.010-0.167 Joules) after the first volume administration then decreased by a median of -0.062 Joules (range of -0.083 to 0.005 Joules) after the second volume administration. EDPVR curves were derived from the pressure-volume loops and differentiated patients with similar starting end-diastolic pressures. **Conclusions** We present a novel method for the acquisition of data to construct pressure-volume loops. Our protocol focuses on rapid volume administration and fast data acquisition with the goal of increasing preload but recording data prior to compensatory changes in afterload. In each patient, administration of 2.5 mL/kg fluid boluses achieved measurable increases in ventricular end-diastolic pressure and ventricular end-diastolic volume. The construction of pressure-volume loops with varying loading may facilitate in-depth assessment of ventricular mechanics in patients with single ventricle heart disease. The variation of preload may allow for the assessment of EDPVR, therefore ventricular stiffness, and to some extent also the contractile response in such a physiology-modifying situation. This work was submitted to the CHOP 2020 conference. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.10. Computational quantification of patient specific changes in ventricular dynamics associated with pulmonary hypertension

Participant: Martin Genet.

Pulmonary arterial hypertension (PAH) causes an increase in the mechanical loading imposed on the right ventricle (RV) that results in progressive changes to its mechanics and function. Here, we quantify the mechanical changes associated with PAH by assimilating clinical data consisting of reconstructed three-dimensional geometry, pressure, and volume waveforms, as well as regional strains measured in patients with PAH ($n = 12$) and controls ($n = 6$) within a computational modeling framework of the ventricles. Modeling parameters reflecting regional passive stiffness and load-independent contractility as indexed by the tissue active tension were optimized so that simulation results matched the measurements. The optimized parameters were compared with clinical metrics to find usable indicators associated with the underlying mechanical changes. Peak contractility of the RV free wall (RVFW) $\gamma_{RVFW,max}$ was found to be strongly correlated and had an inverse relationship with the RV and left ventricle (LV) end-diastolic volume ratio (i.e., RVEDV/LVEDV) (RVEDV/LVEDV)+0.44, $R^2 = 0.77$). Correlation with RV ejection fraction ($R^2 = 0.50$) and end-diastolic volume index ($R^2 = 0.40$) were comparatively weaker. Patients with $RVEDV/LVEDV > 1.5$ had 25% lower $\gamma_{RVFW,max}$ ($P < 0.05$) than that of the control. On average, RVFW passive stiffness progressively increased with the degree of remodeling as indexed by RVEDV/LVEDV. These results suggest a mechanical basis of using RVEDV/LVEDV as a clinical index for delineating disease severity and estimating RVFW contractility in patients with PAH. See [25] for more detail.

6.5.11. Validation of equilibrated warping-image registration with mechanical regularization-on 3D ultrasound images

Participant: Martin Genet.

Image registration plays a very important role in quantifying cardiac motion from medical images, which has significant implications in the diagnosis of cardiac diseases and the development of personalized cardiac computational models. Many approaches have been proposed to solve the image registration problem; however, due to the intrinsic ill-posedness of the image registration problem, all these registration techniques, regardless of their variabilities, require some sort of regularization. An efficient regularization approach was recently proposed based on the equilibrium gap principle, named equilibrated warping. Compared to previous work, it has been formulated at the continuous level within the finite strain hyperelasticity framework and solved using the finite element method. Regularizing the image registration problem using this principle is advantageous as it produces a realistic solution that is close to that of an hyperelastic body in equilibrium with arbitrary boundary tractions, but no body load.The equilibrated warping method has already been extensively validated on both tagged and untagged magnetic resonance images. In this paper, we provide full validation of the method on 3D ultrasound images, based on the 2011 MICCAI Motion Tracking Challenge data. See [39] for more detail.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Technical contract with CEA-LIST on the modelling of rough interfaces in the context of wave scattering (10k€)

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

ANR JCJC LungManyScale, M. Genet, P. Moireau, D. Chapelle (383 k€) – The lungs' architecture and function are well characterized; however, many fundamental questions remain (e.g., there is no quantitative link between tissue- and organ-level material responses), which represent real health challenges (e.g., Idiopathic Pulmonary Fibrosis is a poorly understood disease, for which a mechanical vicious cycle has been hypothesized, but not demonstrated). The general objective of this project is twofold: (i) scientifically, to better understand pulmonary mechanics, from the alveola to the organ in health and disease; (ii) clinically, to improve diagnosis and prognosis of patients through personalized computational modeling. More precisely, This project aims at developing a many-scale model of the pulmonary biomechanics, linked by computational nonlinear homogenization. The model will integrate the experimental and clinical data produced by partners, through an estimation pipeline that will represent augmented diagnosis and prognosis tools for the clinicians.

ANR ODISSE, P. Moireau, S. Imperiale (154 k€) – Motivated by some recent developments from two different fields of research, that is, observer design for finite-dimensional systems and inverse problems analysis for some PDE systems, the ODISSE project aims at developing rigorous methodological tools for the design of estimating algorithms for infinite-dimensional systems arising from hyperbolic PDE systems.

ANR SIMR, P. Moireau, D. Chapelle (97 k€) SIMR is a multi-disciplinary project seeking a better understanding of the biophysical mechanisms involved in mitral valve (MV) regurgitation diseases, to improve decision-making in patients by helping to determine the optimal timing for surgery. This project aims at facing this major issue with the following main two objectives: (1) Evaluate the biophysical consequences of MV repair and (2) Design numerical tools, for cardiac hemodynamics, fluid-structure interaction and myocardium biomechanics to provide an in silico counterpart of the in vivo data obtained by tension measurement and imaging.

8.1.2. Other funding

IPM-MS project (for Imagerie Polarimétrique de Mueller pour la réalisation d'un système original de caractérisation des propriétés mécaniques des Matériaux Structurés), J.M. Allain (50k€ funded by the LABEX Lasips) – This project, which involves the LPICM laboratory (Ecole Polytechnique, CNRS), the LMS (Ecole Polytechnique, CNRS, Mines ParisTech) and the Centre des Matériaux (Mines ParisTech), aims at developing an optical tool to study the link between the mechanical properties of a material and its hierarchical organization. Despite the development of new methods to observe the microstructure, one of the limitations is the number of observations that can be obtained on a given sample in a realistic experimental time. To overcome this difficulty, we are planning to use the Mueller polarimetry to obtain at a fast rate (a few frames per second, compared to a few frames per half-hour) relevant information on the local anisotropy of biological (heart, skin) and composite (short fibers composite) samples.

8.2. European Initiatives

8.2.1. Collaborations with Major European Organizations

Partner 1: Division of Biomedical Engineering & Imaging Sciences (BMEIS), St Thomas' Hospital, King's College London, UK

Clinical-modeling topics mostly encompassing congenital heart diseases (BMEIS) acts as “Other participant” in the Inria Associate team ToFMOD, and R. Chabiniok additionally performs clinical MRI exams at St Thomas’ hospital 0.5 days / week.

Partner 2: Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic

Model-constrained image registrations, trans-valvular flow in pathological valves.

Partner 3: Institute for Clinical and Experimental Medicine in Prague

Cardiovascular MRI.

8.3. International Initiatives

8.3.1. Inria Associate Teams Not Involved in an Inria International Labs

8.3.1.1. ToFMod

Title: Cardiac Biomechanical Modeling of Chronic Right Ventricular Loading

International Partner (Institution - Laboratory - Researcher):

UT Southwestern Medical Center, Dallas,Texas (United States), Mohammad Tarique Hussain

Start year: 2018

See also: https://m3disim.saclay.inria.fr/associated-team/

This collaboration aims at addressing a crucial issue in cardiology of congenital heart diseases, namely, the optimal timing of pulmonary valve replacement (PVR) in patients with surgically repaired tetralogy of Fallot (ToF) prone to chronic pulmonary regurgitation or right ventricular outflow tract stenosis. Our strategy consists in exploiting the predictive power of biomechanical modeling to shed light in the decision process. We will start by a detailed proof-of-concept study, based on datasets that will be acquired in patients indicated for percutaneous PVR, prior to the procedure, and in the follow-up at 3- and 12-months post-PVR. These datasets will be first used to calibrate the Inria M3DISIM patient-specific heart model simulating a cardiac cycle (at each follow-up time point) to access the myocardial properties – namely, the active contractility and passive stiffness. The instantaneous tissue properties will be statistically analyzed and compared with the level of reverse remodeling – i.e. the positive outcome of PVR. Secondly, the data at each time point will be used to calibrate and further develop the models of long-term tissue remodeling created by the M3DISIM researchers. It is only by combining such invaluable longitudinal data with biomechanical modeling expertise that progress can be achieved in the above objective, indeed.

8.4. International Research Visitors

8.4.1. Invited researchers

- T. Hussain, A. Tandon (Senior researchers at UTSW Medical Center Dallas): joint work in the scope of the Inria Associate team ToFMOD
- F. Regazzoni (3rd year PhD student from MOX, Milan, Italy): From January until March 2019 and from December 2019, joint work on model learning and data assimilation coupling.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

F. Clément, Session *Biomathématiques, Bioinformatique et Biophysique pour la reproduction*, ReproSciences 2019 Toulouse (France), April 24-26

D. Chapelle, M. Genet: Session co-chairs at the 6th International Conference on Computational and Mathematical Biomedical Engineering (CMBE2019), Sendai City (Japan), June 10-12

9.1.1.2. Member of organizing committees

J.M. Allain, Member of the workshop Biomechanics from cells to tissues organising committee

D. Chapelle, Chair of the organizing committee of the VPH2020 conference (Paris, August 2020)

F. Clément, ReproSciences 2019, Toulouse (France), April 24-26

M. Genet, D. Chapelle, Co-organizers of the Paris-Saclay University Biomechanics Seminar Series (Until June 2019)

P. Moireau, Member of the organizing committee of the VPH2020 conference (Paris, August 2020)

9.1.2. Scientific events selection

9.1.2.1. Reviewer

J.M. Allain, Member of the European Society of Biomechanics conference scientific committee

R. Chabiniok, reviewer for FIMH2019

M. Genet, reviewer for the 10th international conference on Functional Imaging and Modeling of the Heart (FIMH, June 6-8, 2019, Bordeaux)

9.1.3. Journal

9.1.3.1. Member of editorial boards

D. Chapelle, Member of the editorial board of journal *Computers & Structures*

D. Chapelle, Member of the editorial board of journal *ESAIM: M2AN*

F. Clément *Frontiers*, review editor (Systems Endocrinology)

P. Le Tallec, Member of the editorial board of journal *Computer Methods in Applied Mechanics and Engineering*

P. Le Tallec, Member of the editorial board of journal *Computer and Structures*

9.1.3.2. Reviewer - Reviewing activities

J.M. Allain, reviewer for “Acta Biomaterialia”, “Journal of Anatomy”, “Journal of Biomechanics” and “Journal of the Mechanical Behavior of Biomedical Materials”

R. Chabiniok, reviewer for “Philosophical Transactions of the Royal Society A”

F. Clément, reviewer for “Endocrinology”, “PloS Comp. Biol.”, “Appl. Math. Mod.”

M. Genet, reviewer for “Acta Biomaterialia” and “Inverse Problems in Science & Engineering”

D. Chapelle, reviewer for “Biomechanics and Modeling in Mechanobiology”, “Computers & Structures”, “International Journal for Numerical Methods in Biomedical Engineering”

S. Imperiale, reviewer for “Numerische Mathematik” and “Proceeding of the royal society A”

P. Moireau, reviewer for “Journal of Computational Physics”, “Biomechanics and Modeling in Mechanobiology”

9.1.4. Invited talks

F. Clément. Invited seminar at MaiAGE (France), invited mini-symposium talk to Equadiff, Leiden, The Netherlands

M. Genet, invited keynote at the 90th Annual Meeting of the International Association of Applied Mathematics and Mechanics (GAMM, February 18-22), Vienna, Austria

M. Genet, invited keynote at the francilian mechanics meeting (RFM, Mai 27-28), Fontainebleau

M. Genet, invited keynote at the “Biomechanics from cell to tissue” workshop (November 8), Gif-sur-Yvette

P. Le Tallec, invited speaker at at the 13th International Conference on Advanced Computational Engineering and Experimenting (ACEX, July 1-5), Athens, Greece

D. Chapelle, invited lecturer at iHeart workshop (22-24 July), Varese, Italy

P. Moireau, invited lecture at the GDR MAMOVI 2019 days, Tours, France

P. Moireau, invited lecture at BME symposium 2019, Ecole Polytechnique, France

P. Moireau, invited seminars at LMA Marseille France and CMAP Polytechnique, France

9.1.5. Leadership within the scientific community

D. Chapelle, Member of the board of directors of the VPH Institute

D. Chapelle, Member of the steering committee of the BioMedical Engineering (BME) Institute coordinated by Ecole Polytechnique

F. Clément, member of the direction and scientific board of GdR REPRO (Integrative and translational approaches of human and animal reproduction)

F. Clément, expert of the BCDE (Cell Biology, Development and Evolution) ITMO (Multi OrganizationThematic Institute) of the French National Alliance for Life and Health Sciences Aviesan.

P. Moireau, Member of the steering committee of Department of Mathematics of Université Paris Saclay and Jacques Hadamard Foundation

9.1.6. Scientific expertise

J.M. Allain, Reviewer for the NSERC (Canadian ANR)

R. Chabiniok, Reviewer for Swiss National Science Foundation

F. Clément, member of the INRA DR2 admissibility and admission juries “Agronomie, biologie et amélioration des plantes, sciences du numérique, sciences économiques et sociales”

M. Genet, Reviewer for the Swedish Foundation for Strategic Research

S. Imperiale, Consultant for CEA

P. Moireau, Reviewer for ISCD (Data science Institute, Paris Sorbonne Université)

P. Moireau, Reviewer for ANR and Member of the evaluation committee 46 "Numercial models, simulations, applications"

P. Moireau, member of the MAP5 Paris Descartes, MCF (assistant professor) admission jury

9.1.7. Research administration

J.M. Allain, Scientific Advisory Board, chair BioMecAM

R. Chabiniok, in charge of the objectives of Inria Associate team ToFMOD (with UT Southwestern Medical center Dallas, USA)

D. Chapelle, Head of Science of Inria Saclay-Ile-de-France, and member of the Inria Evaluation Committee

P. Moireau, Member of the LMS board of direction

P. Le Tallec, Director of the Laboratory of Solid Mechanics at Ecole Polytechnique

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Bachelor: J.-M. Allain, Supervision of the introductory projects in physics, 15h, B1, Ecole Polytechnique, France

Bachelor: J.-M. Allain, “Classical mechanics”, 24h, B2, Ecole Polytechnique, France

Bachelor: J.-M. Allain, “Advanced labwork”, 12h, B3, Ecole Polytechnique, France

Bachelor: M. Genet, “Continuum Mechanics I”, B3, 40h, École Polytechnique, France

Bachelor: M. Genet, “Continuum Mechanics II”, B3, 40h, École Polytechnique, France

Bachelor: S. Imperiale, “MA102 – Analyse pour les EDP”, 24h, B3, ENSTA ParisTech, France

Bachelor: P. Le Tallec , “Mécanique des Milieux Continus 2. ”, 24h, L3, École Polytechnique, France

Bachelor: P. Le Tallec , “Mechanics of Continuous Media and Structure. ”, 24h, Y4, Shanghai Jiao Tong Elite Institute of Technology, China

Master: J.-M. Allain, “Statistical mechanics: application to cell motility”, 20h, M2, Ecole Polytechnique, France

Master: D. Chapelle, “Biomechanical Modeling of Active Tissues”, 33h, M2, Université Paris-Saclay, France

Master: M. Genet, “Plasticity and Fracture”, 18h, M1, Ecole Polytechnique, France

Master: P. Le Tallec, “Solid and Continuum Mechanics”, 12h, M1, Master of Nuclear Energy, Université Paris-Saclay, France

Master: P. Moireau, “Biomechanical Modeling of Active Tissues”, 8h, M2, Université Paris-Saclay, France

Master: P. Moireau, “Méthodes et problèmes inverses en dynamique des populations”, 16h, M2, UPMC, France

Master: P. Moireau, “AMS305 – Complétion de données et identification dans les problèmes gouvernés par des équations aux dérivées partielles”, 16h, M2, Université Paris-Saclay, France

9.2.2. Supervision

PhD: F. Caforio, “Modélisation mathématique et numérique de la propagation d’ondes élastique dans le coeur”, supervisors: D. Chapelle and S. Imperiale, defended in January

PhD: F. Kimmig, “Multi-scale modeling of muscle contraction – From stochastic dynamics of molecular motors to continuum mechanics, in interaction with experimental assays”, supervisors: M. Caruel and D. Chapelle, defended in December

PhD: F. Robin “Multiscale modeling of the morphodynamics in ovarian follicles”, supervisors: F. Clément & Romain Yvinec [INRA], defended in September

PhD: F. Wijanto, “Multiscale Mechanics of soft tissues”, supervisors: J.-M. Allain and M. Caruel, defended in December

PhD in progress: G. Ballif “Stochastic multiscale modeling in developmental and reproductive biology”, started October 2019, supervisors: F. Clément & R. Yvinec [INRA]

PhD in progress: J. Manganotti, “Energy preserving cardiac circulation models”, started october 2019, supervisors: S. Imperiale and P. Moireau,

PhD in progress: N. Tueni, “Multiscale properties of the passive cardiac muscle”, started 01/2018, supervisors: J.M. Allain and M. Genet

PhD in progress: C. Giraudet, “Cornea biomechanics”, started 10/2018; supervisors: J.M. Allain and P. Le Tallec

PhD in progress: E. Berberoglu (ETHZ, Switzerland), “Image Guided Computational Cardiac Mechanics”, started February 2017, supervisors: M. Genet and S. Kozerke (ETHZ, Switzerland)

PhD in progress: C. Della Valle, “Modélisation et estimation des dynamiques d’assemblage de proteines”, supervisors: M. Doumic and P. Moireau

PhD in progress: M. Gusseva, “Cardiac Biomechanical Modeling for Chronic Ventricular Loading”, supervisors: R. Chabiniok, D. Chapelle, T. Hussain, started in December 2017

PhD in progress: A. Le Gall, “Cardiac modelling for monitoring purposes during general anaesthesia and at Intensive Care Unit”, supervisors: R. Chabiniok, D. Chapelle, E. Gayat, started in November 2016

PhD in progress: C. Patte, “Lung multiscale poromechanical modeling: from breathing to pulmonary fibrosis-induced chronic remodeling”, supervisors: M. Genet and D. Chapelle, started November 2017

PhD in progress: J. Joachim “Développement d’une nouvelle méthode d’administration automatisée des médicaments utilisés chez les patients sous anesthésie générale basée sur un monitorage totalement non-invasif”, supervisor: E. Gayat, started September 2018

9.2.3. Juries

J.M. Allain, PhD Jury of T. Cochereau (referee), Grenoble University, PhD Advisor L. Bailly, March 18

J.M. Allain, PhD Jury of T. Rongsawat (referee), Montpellier University, PhD Advisor H. Sentenac, December 13

D. Chapelle, PhD Jury of M. Pfaller (referee), TUM, PhD Advisor W. Wall, April 26

D. Chapelle, PhD Jury of F. Vallée, PhD Advisor E. Gayat, Sorbonne Université, June 20

D. Chapelle, PhD Jury of T. Boucneau, PhD Advisor X. Maître, Université Paris-Saclay, July 3

F. Clément, PhD Jury of H. Martin, Sorbonne Université, July 15

F. Clément, PhD Jury of A. Perrillat-Mercerot, Université de Poitiers, October 22

P. Moireau, PhD Jury of Lorenzo Sala, PhD Advisor C. Prud’homme, Université de Strasbourg, September 27

P. Moireau, PhD Jury of Sebastian Reyes Riffo, PhD Advisor J. Salomon, Université Paris Dauphine, November 29

P. Moireau, PhD Jury of François Kimmig, PhD Advisor D. Chapelle, Institut Polytechnique de Paris, December 6

9.3. Popularization

C. Patte, J. Diaz and J. Manganotti, leading of a scientific activity for the Fête des Sciences day organized by Inria. October 11

C. Patte, creation of a scientific activity for middle and high school students as part of a popularization doctoral mission

C. Patte, supervision of a research activity for the RJMI (Rendez-vous des Jeunes Mathématiciennes et Informaticiennes) organized by Inria. October 21 & 22

10. Bibliography

Major publications by the team in recent years

[1] J. Albella Martínez, S. Imperiale, P. Joly, J. Rodríguez.*Solving 2D linear isotropic elastodynamics by means of scalar potentials: a new challenge for finite elements*, in "Journal of Scientific Computing", 2018 [*DOI :* 10.1007/s10915-018-0768-9], https://hal.inria.fr/hal-01803536

[2] M. Caruel, P. Moireau, D. Chapelle.*Stochastic modeling of chemical-mechanical coupling in striated muscles*, in "Biomechanics and Modeling in Mechanobiology", 2018, forthcoming [*DOI :* 10.1007/s10237-018-1102-z], https://hal.inria.fr/hal-01928279

[3] R. CHABINIOK, P. MOIREAU, P.-F. LESAULT, A. RAHMOUNI, J.-F. DEUX, D. CHAPELLE.*Estimation of tissue contractility from cardiac cine-MRI using a biomechanical heart model*, in "Biomechanics and Modeling in Mechanobiology", 2012, vol. 11, n^{o} 5, p. 609-630 [*DOI :* 10.1007/S10237-011-0337-8], http://hal.inria.fr/hal-00654541

[4] D. CHAPELLE, K. BATHE.*The Finite Element Analysis of Shells - Fundamentals - Second Edition*, Computational Fluid and Solid Mechanics, Springer, 2011, 410 [*DOI :* 10.1007/978-3-642-16408-8], http://hal.inria.fr/hal-00654533

[5] D. CHAPELLE, N. CÎNDEA, P. MOIREAU.*Improving convergence in numerical analysis using observers - The wave-like equation case*, in "Mathematical Models and Methods in Applied Sciences", 2012, vol. 22, n^{o} 12 [*DOI :* 10.1142/S0218202512500406], http://hal.inria.fr/inria-00621052

[6] D. CHAPELLE, P. LE TALLEC, P. MOIREAU, M. SORINE.*An energy-preserving muscle tissue model: formulation and compatible discretizations*, in "International Journal for Multiscale Computational Engineering", 2012, vol. 10, n^{o} 2, p. 189-211 [*DOI :* 10.1615/INTJMULTCOMPENG.2011002360], http://hal.inria.fr/hal-00678772

[7] D. CHAPELLE, P. MOIREAU.*General coupling of porous flows and hyperelastic formulations – From thermodynamics principles to energy balance and compatible time schemes*, in "European Journal of Mechanics - B/Fluids", 2014, vol. 46, p. 82-96, Updated version of previously published research report [*DOI :* 10.1016/J.EUROMECHFLU.2014.02.009], https://hal.inria.fr/inria-00520612

[8] A. COLLIN, S. IMPERIALE.*Mathematical analysis and 2-scale convergence of a heterogeneous microscopic bidomain model*, in "Mathematical Models and Methods in Applied Sciences", 2018, forthcoming, https://hal.inria.fr/hal-01759914

[9] B. LYNCH, S. BANCELIN, C. BONOD-BIDAUD, J.-B. GUEUSQUIN, F. RUGGIERO, M.-C. SCHANNE-KLEIN, J.-M. ALLAIN.*A novel microstructural interpretation for the biomechanics of mouse skin derived from multiscale characterization*, in "Acta Biomaterialia", 2017, vol. 50, p. 302-311 [*DOI :* 10.1016/J.ACTBIO.2016.12.051], https://hal.archives-ouvertes.fr/hal-01531321

[10] P. MOIREAU.*A Discrete-time Optimal Filtering Approach for Non-linear Systems as a Stable Discretization of the Mortensen Observer*, in "ESAIM: Control, Optimisation and Calculus of Variations", 2017, forthcoming, https://hal.inria.fr/hal-01671271

[11] M. SERMESANT, R. CHABINIOK, P. CHINCHAPATNAM, T. MANSI, F. BILLET, P. MOIREAU, J.-M. PEYRAT, K. C. WONG, J. RELAN, K. S. RHODE, M. GINKS, P. LAMBIASE, H. DELINGETTE, M. SORINE, C. A. RINALDI, D. CHAPELLE, R. RAZAVI, N. AYACHE.*Patient-Specific Electromechanical Models of the Heart for Prediction of the Acute Effects of Pacing in CRT: a First Validation*, in "Medical Image Analysis", 2012, vol. 16, n^{o} 1, p. 201-215 [*DOI :* 10.1016/J.MEDIA.2011.07.003], http://hal.inria.fr/inria-00616191

Publications of the year

Doctoral Dissertations and Habilitation Theses

[12] F. CAFORIO.*Mathematical modelling and numerical simulation of elastic wave propagation in soft tissues with application to cardiac elastography*, Université Paris-Saclay, January 2019, https://pastel.archives-ouvertes.fr/tel-02262282

[13] F. KIMMIG.*Modélisation multi-échelles de la contraction musculaire : De la dynamique stochastique des moteurs moléculaires à la mécanique des milieux continus*, Université Paris-Saclay, December 2019, https://tel.archives-ouvertes.fr/tel-02439432

Articles in International Peer-Reviewed Journal

[14] J. ALBELLA MARTÍNEZ, H. BEN DHIA, S. IMPERIALE, J. RODRÍGUEZ.*Mathematical and numerical study of transient wave scattering by obstacles with the Arlequin Method*, in "SIAM Journal on Numerical Analysis", 2019, forthcoming [*DOI :* 10.1137/19M1263959], https://hal.inria.fr/hal-01898420

[15] D. H. BAFFET, M. J. GROTE, S. IMPERIALE, M. KACHANOVSKA.*Energy Decay and Stability of a Perfectly Matched Layer For the Wave Equation*, in "Journal of Scientific Computing", November 2019 [*DOI :* 10.1007/S10915-019-01089-9], https://hal.archives-ouvertes.fr/hal-01865484

[16] B. BURTSCHELL, P. MOIREAU, D. CHAPELLE.*Numerical analysis for an energy-stable total discretization of a poromechanics model with inf-sup stability*, in "Acta Mathematicae Applicatae Sinica", 2019, vol. 35, n^{o} 1, p. 28-53 [*DOI :* 10.1007/S10255-019-0799-5], https://hal.inria.fr/hal-01928312

[17] F. CAFORIO, S. IMPERIALE.*High-order discrete fourier transform for the solution of the Poisson equation*, in "SIAM Journal on Scientific Computing", June 2019, vol. 41, n^{o} 5, p. A2747-A2771 [*DOI :* 10.1137/18M1225410], https://hal.inria.fr/hal-01914257

[18] M. CARUEL, P. MOIREAU, D. CHAPELLE.*Stochastic modeling of chemical-mechanical coupling in striated muscles*, in "Biomechanics and Modeling in Mechanobiology", 2019 [*DOI :* 10.1007/S10237-018-1102-Z], https://hal.inria.fr/hal-01928279

[19] J. CHABASSIER, J. DIAZ, S. IMPERIALE.*Construction and analysis of fourth order, energy consistent, family of explicit time discretizations for dissipative linear wave equations*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2019, forthcoming, https://hal.inria.fr/hal-01894238

[20] F. CLÉMENT, B. LAROCHE, F. ROBIN.*Analysis and numerical simulation of an inverse problem for a structured cell population dynamics model*, in "Mathematical Biosciences & Engineering", 2019, vol. 16, n^{o} 4, p. 3018-3046 [*DOI :* 10.3934/MBE.2019150], https://hal.archives-ouvertes.fr/hal-02154588

[21] F. CLÉMENT, F. ROBIN, R. YVINEC.*Analysis and calibration of a linear model for structured cell populations with unidirectional motion : Application to the morphogenesis of ovarian follicles*, in "SIAM Journal on Applied Mathematics", February 2019, vol. 79, n^{o} 1, p. 207-229, https://arxiv.org/abs/1712.05372 [*DOI :* 10.1137/17M1161336], https://hal.archives-ouvertes.fr/hal-01852560

[22] A. COLLIN, S. IMPERIALE, P. MOIREAU, J.-F. GERBEAU, D. CHAPELLE.*Apprehending the effects of mechanical deformations in cardiac electrophysiology - An homogenization approach*, in "Mathematical Models and Methods in Applied Sciences", December 2019, vol. 29, n^{o} 13, p. 2377-2417 [*DOI :* 10.1142/S0218202519500490], https://hal.inria.fr/hal-02267941

[23] J. DIANI, P. LE TALLEC.*A fully equilibrated microsphere model with damage for rubberlike materials*, in "Journal of the Mechanics and Physics of Solids", March 2019, vol. 124, p. 702-713 [*DOI :* 10.1016/J.JMPS.2018.11.021], https://hal.archives-ouvertes.fr/hal-01986861

[24] G. DUCOURTHIAL, J. AFFAGARD, M. SCHMELTZ, X. SOLINAS, M. LOPEZ-PONCELAS, C. BONOD-BIDAUD, R. RUBIO-AMADOR, F. RUGGIERO, J.-M. ALLAIN, E. BEAUREPAIRE, M.-C. SCHANNE-KLEIN.*Monitoring dynamic collagen reorganization during skin stretching with fast polarization-resolved second harmonic generation imaging*, in "Journal of Biophotonics", January 2019, vol. 12, n^{o} 5, e201800336, https://arxiv.org/abs/1904.09177 [*DOI :* 10.1002/JBIO.201800336], https://hal.archives-ouvertes.fr/hal-02103594

[25] H. N. T. FINSBERG, J. SUNDNES, C. XI, L. C. LEE, X. ZHAO, J. L. TAN, M. GENET, L. ZHONG, S. T. WALL.*Computational quantification of patient specific changes in ventricular dynamics associated with pulmonary hypertension*, in "AJP - Heart and Circulatory Physiology", 2019 [*DOI :* 10.1152/AJPHEART.00094.2019], https://hal.archives-ouvertes.fr/hal-02375749

[26] J.-M. FRACHISSE, S. THOMINE, J.-M. ALLAIN.*Calcium and plasma membrane force-gated ion channels behind development*, in "Current Opinion in Plant Biology", 2019, vol. 53, p. 57–64 [*DOI :* 10.1016/J.PBI.2019.10.006], https://hal.archives-ouvertes.fr/hal-02389243

[27] M. GENET.*A relaxed growth modeling framework for controlling growth-induced residual stresses*, in "Clinical Biomechanics", 2019 [*DOI :* 10.1016/J.CLINBIOMECH.2019.08.015], https://hal.archives-ouvertes.fr/hal-02069113

[28] M. GUICHARD, J.-M. ALLAIN, M. W. BIANCHI, J.-M. FRACHISSE.*Root Hair Sizer: an algorithm for high throughput recovery of different root hair and root developmental parameters*, in "Plant Methods", December 2019, vol. 15, n^{o} 1 [*DOI :* 10.1186/S13007-019-0483-Z], https://hal.archives-ouvertes.fr/hal-02293640

[29] S. IMPERIALE, P. MOIREAU, A. TONNOIR.*Analysis of an observer strategy for initial state reconstruction of wave-like systems in unbounded domains*, in "ESAIM: Control, Optimisation and Calculus of Variations", April 2019 [*DOI :* 10.1051/COCV/2019026], https://hal.inria.fr/hal-01898440

[30] F. KIMMIG, D. CHAPELLE, P. MOIREAU.*Thermodynamic properties of muscle contraction models and associated discrete-time principles*, in "Advanced Modeling and Simulation in Engineering Sciences", 2019, forthcoming [*DOI :* 10.1186/S40323-019-0128-9], https://hal.inria.fr/hal-02065957

[31] Z. MA, H. MAITOURNAM, P. LE TALLEC.*An energy-based strategy for fatigue analysis in presence of general multi-axial time varying loadings*, in "International Journal of Fatigue", March 2020, vol. 132, 105367 [*DOI :* 10.1016/J.IJFATIGUE.2019.105367], https://hal.archives-ouvertes.fr/hal-02421856

[32] A. MARBOEUF, A. CLAISSE, P. LE TALLEC.*Conservative and entropy controlled remap for multi-material ALE simulations with space-staggered schemes*, in "Journal of Computational Physics", August 2019, vol. 390, p. 66-92 [*DOI :* 10.1016/J.JCP.2019.04.017], https://hal.archives-ouvertes.fr/hal-02421917

[33] M. POSTEL, A. KARAM, G. PÉZERON, S. SCHNEIDER-MAUNOURY, F. CLÉMENT.*A multiscale mathematical model of cell dynamics during neurogenesis in the mouse cerebral cortex*, in "BMC Bioinformatics", December 2019, vol. 20, n^{o} 1 [*DOI :* 10.1186/S12859-019-3018-8], https://hal.inria.fr/hal-01783141

[34] C. ZHANG, A. COLLIN, P. MOIREAU, A. TROUVÉ, M. ROCHOUX.*Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment*, in "Proceedings of the Combustion Institute", 2019, vol. 37, n^{o} 3, p. 4201-4209 [*DOI :* 10.1016/J.PROCI.2018.07.112], https://hal.inria.fr/hal-01969130

International Conferences with Proceedings

[35] E. BERBEROGLU, C. T. STOECK, P. MOIREAU, S. KOZERKE, M. GENET.*Validation of Finite Element Image Registration-based Cardiac Strain Estimation from Magnetic Resonance Images*, in "GAMM 2019 - 90th Annual Meeting of the International Association of Applied Mathematics and Mechanics", Vienna, Austria, PAMM proceedings in applied mathematics and mechanics, February 2019 [*DOI :* 10.1002/PAMM.201900418], https://hal.archives-ouvertes.fr/hal-02375735

[36] R. CHABINIOK, G. BUREAU, A. GROTH, J. TINTERA, J. WEESE, D. CHAPELLE, P. MOIREAU.*Cardiac displacement tracking with data assimilation combining a biomechanical model and an automatic contour detection*, in "FIMH 2019 - Proceedings of Functional Imaging and Modeling of Heart", Bordeaux, France, Lecture Notes in Computer Science (LNCS), June 2019, vol. 11504, p. 405-414, https://hal.archives-ouvertes.fr/hal-02186804

[37] A. GÉRARD, A. COLLIN, G. BUREAU, P. MOIREAU, Y. COUDIÈRE.*Model assessment through data assimilation of realistic data in cardiac electrophysiology*, in "FIMH 2019 - 10th Functional Imaging and Modeling of the Heart", Bordeaux, France, June 2019, https://hal.inria.fr/hal-02172102

[38] A. LE GALL, F. VALLÉE, D. CHAPELLE, R. CHABINIOK.*Minimally-invasive estimation of patient-specific end-systolic elastance using a biomechanical heart model*, in "FIMH 2019 - 10th Functional Imaging and Modelling of the Heart", Bordeaux, France, June 2019, https://hal.archives-ouvertes.fr/hal-02153611

[39] L. C. LEE, M. GENET.*Validation of Equilibrated Warping-image registration with mechanical regularization-on 3D ultrasound images*, in "FIMH 2019 - International Conference on Functional Imaging and Modeling of the Heart", Bordeaux, France, Lecture notes in computer science, Springer, May 2019, vol. 11504, p. 334-341 [*DOI :* 10.1007/978-3-030-21949-9_36], https://hal.archives-ouvertes.fr/hal-02146580

[40] B. RUIJSINK, K. ZUGAJ, K. PUSHPARAJAH, R. CHABINIOK.*Model-based indices of early-stage cardiovascular failure and its therapeutic management in Fontan patients*, in "FIMH 2019 - 10th Functional Imaging and Modeling of Heart", Bordeaux, France, Lecture Notes in Computer Science (LNCS), June 2019, vol. 11504, p. 379-387, https://hal.archives-ouvertes.fr/hal-02186819

Conferences without Proceedings

[41] G. BECK, S. IMPERIALE, P. JOLY.*Effective models for non-perfectly conducting thin coaxial cables*, in "Waves 2019 - 14th International Conference on Mathematical and Numerical Aspects of Wave Propagation", Vienna, Austria, 2019, https://hal.archives-ouvertes.fr/hal-02414849

[42] M. GENET.*Medical images registration with finite elements and mechanical regularization*, in "GAMM 2019 - 90th Annual Meeting of the International Association of Applied Mathematics and Mechanics", Vienna, Austria, February 2019, https://hal.archives-ouvertes.fr/hal-02069100

[43] C. PATTE, M. GENET, C. FETITA, P.-Y. BRILLET, D. CHAPELLE.*Mécanique pulmonaire personnalisée : modélisation et estimation - Application à la fibrose pulmonaire*, in "CSMA 2019 - 14ème Colloque National en Calcul des Structures", Giens, France, May 2019, https://hal.archives-ouvertes.fr/hal-02068920

[44] K. ŠKARDOVÁ, M. RAMBAUSEK, R. CHABINIOK, M. GENET.*Mechanical and Imaging Models-based Image Registration*, in "VipIMAGE 2019 - VII ECCOMAS Thematic Conference on Computational Vision and

Medical Image Processing", Porto, Portugal, 2019 [*DOI :* 10.1007/978-3-030-32040-9_9], https://hal.archives-ouvertes.fr/hal-02186828

Scientific Books (or Scientific Book chapters)

[45] J.-M. ALLAIN, B. LYNCH, M.-C. SCHANNE-KLEIN.*Multiscale characterization of skin mechanics through in situ imaging*, in "Skin Biophysics : From Experimental Characterisation to Advanced Modelling", G. LIMBERT (editor), Studies in Mechanobiology, Tissue Engineering and Biomaterials, Springer International Publishing, 2019, vol. 22, p. 265-280 [*DOI :* 10.1007/978-3-030-13279-8_8], https://hal-polytechnique.archives-ouvertes.fr/hal-02114148

[46] S. BANCELIN, B. LYNCH, C. BONOD-BIDAUD, P. DOKLÁDAL, F. RUGGIERO, J.-M. ALLAIN, M.-C. SCHANNE-KLEIN.*Combination of Traction Assays and Multiphoton Imaging to Quantify Skin Biomechanics*, in "Collagen : Methods and Protocols", I. SAGI, N. A. AFRATIS (editors), Methods in Molecular Biology, Humana Press, March 2019, vol. 1944, p. 145-155 [*DOI :* 10.1007/978-1-4939-9095-5_11], https://hal-polytechnique.archives-ouvertes.fr/hal-02115604

Research Reports

[47] Y. MOCQUARD, F. ROBIN, B. SERICOLA, E. ANCEAUME.*Average-based Population Protocols : Explicit and Tight Bounds of the Convergence Time*, Irisa, July 2019, https://hal.archives-ouvertes.fr/hal-02178618

Patents and standards

[48] D. CHAPELLE, P. MOIREAU, M. PERNOT, M. TANTER.*Dispositif cardiaque*, 2019, n^{o} 1908859, https://hal.inria.fr/hal-02262214

Other Publications

[49] J. ALBELLA MARTÍNEZ, S. IMPERIALE, P. JOLY, J. RODRÍGUEZ.*Numerical Analysis of a Method for Solving 2D Linear Isotropic Elastodynamics with Free Boundary Condition using Potentials and Finite Elements*, November 2019, working paper or preprint [*DOI :* 10.1007/s10915-018-0768-9], https://hal.inria.fr/hal-02345808

[50] C. BONNET, K. CHAHOUR, F. CLÉMENT, M. POSTEL, R. YVINEC.*Multiscale population dynamics in reproductive biology: singular perturbation reduction in deterministic and stochastic models*, March 2019, https://arxiv.org/abs/1903.08555 - working paper or preprint, https://hal.inria.fr/hal-02075253

[51] F. CAFORIO, S. IMPERIALE.*Mathematical analysis of a penalization strategy for incompressible elastodynamics*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02389286

[52] F. CAFORIO, S. IMPERIALE.*Mathematical modelling of transient shear wave elastography in the heart*, March 2019, working paper or preprint, https://hal.inria.fr/hal-01992069

[53] J. CHABASSIER, S. IMPERIALE.*Construction and convergence analysis of conservative second order local time discretisation for wave equations based on domain decomposition*, August 2019, working paper or preprint, https://hal.inria.fr/hal-01894357

[54] F. CLÉMENT, F. ROBIN, R. YVINEC.*Stochastic nonlinear model for somatic cell population dynamics during ovarian follicle activation*, March 2019, https://arxiv.org/abs/1903.01316 - working paper or preprint, https://hal.inria.fr/hal-02057983

Project-Team MEXICO

Modeling and Exploitation of Interaction and Concurrency

IN COLLABORATION WITH: Laboratoire specification et vérification (LSV)

IN PARTNERSHIP WITH:
CNRS

Ecole normale supérieure de Cachan

RESEARCH CENTER
Saclay - Île-de-France

THEME
Proofs and Verification

Table of contents

Project-Team MEXICO

Creation of the Team: 2009 March 01, updated into Project-Team: 2011 January 01

Keywords:

Computer Science and Digital Science:

A2.3. - Embedded and cyber-physical systems
A2.3.2. - Cyber-physical systems
A2.3.3. - Real-time systems
A2.4.1. - Analysis
A2.4.2. - Model-checking
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A7.1. - Algorithms
A7.1.1. - Distributed algorithms
A7.2. - Logic in Computer Science
A7.3.1. - Computational models and calculability
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.7. - Graph theory
A8.8. - Network science
A8.9. - Performance evaluation
A8.11. - Game Theory

Other Research Topics and Application Domains:

B1.1.2. - Molecular and cellular biology
B1.1.7. - Bioinformatics
B1.1.10. - Systems and synthetic biology
B7.1. - Traffic management
B7.2.1. - Smart vehicles

1. Team, Visitors, External Collaborators

Research Scientists

Stefan Haar [Team leader, Inria, Senior Researcher, HDR]
Laurent Fribourg [CNRS, Senior Researcher, until Feb 2019, HDR]
Matthias Fuegger [CNRS, Researcher]

Faculty Members

Thomas Chatain [École Normale Supérieure de Cachan, Associate Professor, HDR]
Philippe Dague [Univ Paris-Sud, Professor, Délégation until Jun 2019, Invited Researcher from Jul 2019, Emeritus Professor from Sep 2019, HDR]
Serge Haddad [École Normale Supérieure de Cachan, Professor, HDR]
Stefan Schwoon [École Normale Supérieure de Cachan, Associate Professor, HDR]
Lina Ye [Centrale-Supélec, Associate Professor, Délégation, from Sep 2019]

External Collaborator

Benoît Barbot [Univ Paris-Est Marne La Vallée]

PhD Students

Mathilde Boltenhagen [CNRS, PhD Student]
Jawher Jerray [Univ Paris-Nord, PhD Student, until Feb 2019]
Igor Khmelnitsky [École Normale Supérieure de Cachan, PhD Student]
Juraj Kolcak [Inria, PhD Student]
Hugues Mandon [Inria, PhD Student, until Sep 2019]
Adnane Saoud [CNRS, PhD Student, until Sep 2019]

Post-Doctoral Fellows

Da Jung Cho [CNRS, Post-Doctoral Fellow, until Apr 2019]
Daniele Zonetti [CNRS, Post-Doctoral Fellow, until Feb 2019]

Administrative Assistants

Adeline Lochet [Inria, Administrative Assistant, until May 2019]
Emmanuelle Perrot [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Scientific Objectives

2.1.1. Introduction.

In the increasingly networked world, reliability of applications becomes ever more critical as the number of users of, e.g., communication systems, web services, transportation etc., grows steadily. Management of networked systems, in a very general sense of the term, therefore is a crucial task, but also a difficult one.

MExICo strives to take advantage of distribution by orchestrating cooperation between different agents that observe local subsystems, and interact in a localized fashion.

The need for applying formal methods in the analysis and management of complex systems has long been recognized. It is with much less unanimity that the scientific community embraces methods based on asynchronous and distributed models. Centralized and sequential modeling still prevails.

However, we observe that crucial applications have increasing numbers of users, that networks providing services grow fast both in the number of participants and the physical size and degree of spatial distribution. Moreover, traditional *isolated* and *proprietary* software products for local systems are no longer typical for emerging applications.

In contrast to traditional centralized and sequential machinery for which purely functional specifications are efficient, we have to account for applications being provided from diverse and non-coordinated sources. Their distribution (e.g. over the Web) must change the way we verify and manage them. In particular, one cannot ignore the impact of quantitative features such as delays or failure likelihoods on the functionalities of composite services in distributed systems.

We thus identify three main characteristics of complex distributed systems that constitute research challenges:

- *Concurrency* of behavior;
- *Interaction* of diverse and semi-transparent components; and
- management of *Quantitative* aspects of behavior.

2.1.2. Concurrency

The increasing size and the networked nature of communication systems, controls, distributed services, etc. confront us with an ever higher degree of parallelism between local processes. This field of application for our work includes telecommunication systems and composite web services. The challenge is to provide sound theoretical foundations and efficient algorithms for management of such systems, ranging from controller synthesis and fault diagnosis to integration and adaptation. While these tasks have received considerable attention in the *sequential* setting, managing *non-sequential* behavior requires profound modifications for existing approaches, and often the development of new approaches altogether. We see concurrency in distributed systems as an opportunity rather than a nuisance. Our goal is to *exploit* asynchronicity and distribution as an advantage. Clever use of adequate models, in particular *partial order semantics* (ranging from Mazurkiewicz traces to event structures to MSCs) actually helps in practice. In fact, the partial order vision allows us to make causal precedence relations explicit, and to perform diagnosis and test for the dependency between events. This is a conceptual advantage that interleaving-based approaches cannot match. The two key features of our work will be *(i)* the exploitation of concurrency by using asynchronous models with partial order semantics, and *(ii)* distribution of the agents performing management tasks.

2.1.3. Interaction

Systems and services exhibit non-trivial *interaction* between specialized and heterogeneous components. A coordinated interplay of several components is required; this is challenging since each of them has only a limited, partial view of the system's configuration. We refer to this problem as *distributed synthesis* or *distributed control*. An aggravating factor is that the structure of a component might be semi-transparent, which requires a form of *grey box management*.

2.1.4. Quantitative Features

Besides the logical functionalities of programs, the *quantitative* aspects of component behavior and interaction play an increasingly important role.

- *Real-time* properties cannot be neglected even if time is not an explicit functional issue, since transmission delays, parallelism, etc, can lead to time-outs striking, and thus change even the logical course of processes. Again, this phenomenon arises in telecommunications and web services, but also in transport systems.
- In the same contexts, *probabilities* need to be taken into account, for many diverse reasons such as unpredictable functionalities, or because the outcome of a computation may be governed by race conditions.
- Last but not least, constraints on *cost* cannot be ignored, be it in terms of money or any other limited resource, such as memory space or available CPU time.

2.1.5. Evolution and Perspectives

Since the creation of *MExICo*, the weight of *quantitative* aspects in all parts of our activities has grown, be it in terms of the models considered (weighted automata and logics), be it in transforming verification or diagnosis verdict into probabilistic statements (probabilistic diagnosis, statistical model checking), or within the recently started SystemX cooperation on supervision in multi-modal transport systems. This trend is certain to continue over the next couple of years, along with the growing importance of diagnosis and control issues.

In another development, the theory and use of partial order semantics has gained momentum in the past four years, and we intend to further strengthen our efforts and contacts in this domain to further develop and apply partial-order based deduction methods.

When no complete model of the underlying dynamic system is available, the analysis of logs may allow to reconstruct such a model, or at least to infer some properties of interest; this activity, which has emerged over the past 10 years on the international level, is referred to as **process mining**. In this emerging activity, we have contributed to unfolding-based process discovery [CI-146], and the study of process alignments [CI-121, CI-96, CI-83, CI-60, CI-33].

Finally, over the past years *biological* challenges have come to the center of our work, in two different directions:

1. **(Re-)programming in discrete concurrent models.** Cellular regulatory networks exhibit highly complex concurrent behaviours that is influenced by a high number of perturbations such as mutations. We are in particular investigating discrete models, both in the form of boolean networks and of Petri nets, to harness this complexity, and to obtain viable methods for two interconnected and central challenges:
   - find *attractors*, i.e. long-run stable states or sets of states, that indicate possible phenotypes of the organism under study, and
   - determine *reprogramming* strategies that apply perturbations in such a way as to steer the cellâs long-run behaviour into some desired phenotype, or away from an undesired one.
2. **Distributed Algorithms in wild or synthetic biological systems.** Since the arrival of Matthias Fuegger in the team, we also work, on the multi-cell level, with a distributed algorithmsâ view on microbiological systems, both with the goal to model and analyze existing microbiological systems as distributed systems, and to design and implement distributed algorithms in synthesized microbiological systems. Major long-term goals are drug production and medical treatment via synthesized bacterial colonies.

3. Research Program

3.1. Concurrency

Participants: Thomas Chatain, Philippe Dague, Stefan Haar, Serge Haddad, Stefan Schwoon.

Concurrency; Semantics; Automatic Control ; Diagnosis ; Verification

Concurrency: Property of systems allowing some interacting processes to be executed in parallel.

Diagnosis: The process of deducing from a partial observation of a system aspects of the internal states or events of that system; in particular, *fault diagnosis* aims at determining whether or not some non-observable fault event has occurred.

Conformance Testing: Feeding dedicated input into an implemented system IS and deducing, from the resulting output of I, whether I respects a formal specification S.

3.1.1. Introduction

It is well known that, whatever the intended form of analysis or control, a *global* view of the system state leads to overwhelming numbers of states and transitions, thus slowing down algorithms that need to explore the state space. Worse yet, it often blurs the mechanics that are at work rather than exhibiting them. Conversely, respecting concurrency relations avoids exhaustive enumeration of interleavings. It allows us to focus on ‘essential’ properties of non-sequential processes, which are expressible with causal precedence relations. These precedence relations are usually called causal (partial) orders. Concurrency is the explicit absence of such a precedence between actions that do not have to wait for one another. Both causal orders and concurrency are in fact essential elements of a specification. This is especially true when the specification is constructed in a distributed and modular way. Making these ordering relations explicit requires to leave the framework of state/interleaving based semantics. Therefore, we need to develop new dedicated algorithms for tasks such as conformance testing, fault diagnosis, or control for distributed discrete systems. Existing solutions for these problems often rely on centralized sequential models which do not scale up well.

3.1.2. Diagnosis

Participants: Stefan Haar, Serge Haddad, Stefan Schwoon, Philippe Dague, Lina Ye.

Fault Diagnosis for discrete event systems is a crucial task in automatic control. Our focus is on *event oriented* (as opposed to *state oriented*) model-based diagnosis, asking e.g. the following questions:

given a - potentially large - *alarm pattern* formed of observations,

- what are the possible *fault scenarios* in the system that *explain* the pattern ?
- Based on the observations, can we deduce whether or not a certain - invisible - fault has actually occurred ?

Model-based diagnosis starts from a discrete event model of the observed system - or rather, its relevant aspects, such as possible fault propagations, abstracting away other dimensions. From this model, an extraction or unfolding process, guided by the observation, produces recursively the explanation candidates.

In asynchronous partial-order based diagnosis with Petri nets [45], [46], [47], one unfolds the *labelled product* of a Petri net model $\mathcal{N}$ and an observed alarm pattern $\mathcal{A}$, also in Petri net form. We obtain an acyclic net giving partial order representation of the behaviors compatible with the alarm pattern. A recursive online procedure filters out those runs *(configurations)* that explain *exactly*$\mathcal{A}$. The Petri-net based approach generalizes to dynamically evolving topologies, in dynamical systems modeled by graph grammars, see [34]

3.1.2.1. Observability and Diagnosability

Diagnosis algorithms have to operate in contexts with low observability, i.e., in systems where many events are invisible to the supervisor. Checking *observability* and *diagnosability* for the supervised systems is therefore a crucial and non-trivial task in its own right. Analysis of the relational structure of occurrence nets allows us to check whether the system exhibits sufficient visibility to allow diagnosis. Developing efficient methods for both verification of *diagnosability checking* under concurrency, and the *diagnosis* itself for distributed, composite and asynchronous systems, is an important field for *MExICo*. In 2019, a new property, manifestability, weaker than diagnosability (dual in some sense to opacity) has been studied in the context of automata and timed automata.

3.1.2.2. Distribution

Distributed computation of unfoldings allows one to factor the unfolding of the global system into smaller *local* unfoldings, by local supervisors associated with sub-networks and communicating among each other. In [46], [36], elements of a methodology for distributed computation of unfoldings between several supervisors, underwritten by algebraic properties of the category of Petri nets have been developed. Generalizations, in particular to Graph Grammars, are still do be done.

Computing diagnosis in a distributed way is only one aspect of a much vaster topic, that of *distributed diagnosis* (see [43], [49]). In fact, it involves a more abstract and often indirect reasoning to conclude whether or not some given invisible fault has occurred. Combination of local scenarios is in general not sufficient: the global system may have behaviors that do not reveal themselves as faulty (or, dually, non-faulty) on any local supervisor's domain (compare [33], [39]). Rather, the local diagnosers have to join all *information* that is available to them locally, and then deduce collectively further information from the combination of their views. In particular, even the *absence* of fault evidence on all peers may allow to deduce fault occurrence jointly, see [51], [52]. Automatizing such procedures for the supervision and management of distributed and locally monitored asynchronous systems is a long-term goal to which *MExICo* hopes to contribute.

3.1.3. Hybrid Systems

Participants: Philippe Dague, Lina Ye, Serge Haddad.

Hybrid systems constitute a model for cyber-physical systems which integrates continuous-time dynamics (modes) governed by differential equations, and discrete transitions which switch instantaneously from one mode to another. Thanks to their ease of programming, hybrid systems have been integrated to power electronics systems, and more generally in cyber-physical systems. In order to guarantee that such systems meet their specifications, classical methods consist in finitely abstracting the systems by discretization of the (infinite) state space, and deriving automatically the appropriate mode control from the specification using standard graph techniques.

Diagnosability of hybrid systems has also been studied through an abstraction / refinement process in terms of timed automata.

3.1.4. Contextual Nets

Participant: Stefan Schwoon.

Assuring the correctness of concurrent systems is notoriously difficult due to the many unforeseeable ways in which the components may interact and the resulting state-space explosion. A well-established approach to alleviate this problem is to model concurrent systems as Petri nets and analyse their unfoldings, essentially an acyclic version of the Petri net whose simpler structure permits easier analysis [44].

However, Petri nets are inadequate to model concurrent read accesses to the same resource. Such situations often arise naturally, for instance in concurrent databases or in asynchronous circuits. The encoding tricks typically used to model these cases in Petri nets make the unfolding technique inefficient. Contextual nets, which explicitly do model concurrent read accesses, address this problem. Their accurate representation of concurrency makes contextual unfoldings up to exponentially smaller in certain situations. An abstract algorithm for contextual unfoldings was first given in [35]. In recent work, we further studied this subject from a theoretical and practical perspective, allowing us to develop concrete, efficient data structures and algorithms and a tool (Cunf) that improves upon existing state of the art. This work led to the PhD thesis of César Rodríguez in 2014 .

Contextual unfoldings deal well with two sources of state-space explosion: concurrency and shared resources. Recently, we proposed an improved data structure, called *contextual merged processes* (CMP) to deal with a third source of state-space explosion, i.e. sequences of choices. The work on CMP [53] is currently at an abstract level. In the short term, we want to put this work into practice, requiring some theoretical groundwork, as well as programming and experimentation.

Another well-known approach to verifying concurrent systems is *partial-order reduction*, exemplified by the tool SPIN. Although it is known that both partial-order reduction and unfoldings have their respective strengths and weaknesses, we are not aware of any conclusive comparison between the two techniques. Spin comes with a high-level modeling language having an explicit notion of processes, communication channels, and variables. Indeed, the reduction techniques implemented in Spin exploit the specific properties of these features. On the other side, while there exist highly efficient tools for unfoldings, Petri nets are a relatively general low-level formalism, so these techniques do not exploit properties of higher language features. Our work on contextual unfoldings and CMPs represents a first step to make unfoldings exploit richer models. In the long run, we wish raise the unfolding technique to a suitable high-level modelling language and develop appropriate tool support.

3.2. Management of Quantitative Behavior

Participants: Thomas Chatain, Stefan Haar, Serge Haddad.

3.2.1. Introduction

Besides the logical functionalities of programs, the *quantitative* aspects of component behavior and interaction play an increasingly important role.

- *Real-time* properties cannot be neglected even if time is not an explicit functional issue, since transmission delays, parallelism, etc, can lead to time-outs striking, and thus change even the logical course of processes. Again, this phenomenon arises in telecommunications and web services, but also in transport systems.
- In the same contexts, *probabilities* need to be taken into account, for many diverse reasons such as unpredictable functionalities, or because the outcome of a computation may be governed by race conditions.
- Last but not least, constraints on *cost* cannot be ignored, be it in terms of money or any other limited resource, such as memory space or available CPU time.

Traditional mainframe systems were proprietary and (essentially) localized; therefore, impact of delays, unforeseen failures, etc. could be considered under the control of the system manager. It was therefore natural, in verification and control of systems, to focus on *functional* behavior entirely.

With the increase in size of computing system and the growing degree of compositionality and distribution, quantitative factors enter the stage:

- calling remote services and transmitting data over the web creates *delays*;
- remote or non-proprietary components are not "deterministic", in the sense that their behavior is uncertain.

Time and *probability* are thus parameters that management of distributed systems must be able to handle; along with both, the *cost* of operations is often subject to restrictions, or its minimization is at least desired. The mathematical treatment of these features in distributed systems is an important challenge, which *MExICo* is addressing; the following describes our activities concerning probabilistic and timed systems. Note that cost optimization is not a current activity but enters the picture in several intended activities.

3.2.2. Probabilistic distributed Systems

Participants: Stefan Haar, Serge Haddad.

3.2.2.1. Non-sequential probabilistic processes

Practical fault diagnosis requires to select explanations of *maximal likelihood*. For partial-order based diagnosis, this leads therefore to the question what the probability of a given partially ordered execution is. In Benveniste et al. [38], [31], we presented a model of stochastic processes, whose trajectories are partially ordered, based on local branching in Petri net unfoldings; an alternative and complementary model based on Markov fields is developed in [48], which takes a different view on the semantics and overcomes the first model's restrictions on applicability.

Both approaches abstract away from real time progress and randomize choices in *logical* time. On the other hand, the relative speed - and thus, indirectly, the real-time behavior of the system's local processes - are crucial factors determining the outcome of probabilistic choices, even if non-determinism is absent from the system.

In another line of research [40] we have studied the likelihood of occurrence of non-sequential runs under random durations in a stochastic Petri net setting. It remains to better understand the properties of the probability measures thus obtained, to relate them with the models in logical time, and exploit them e.g. in *diagnosis*.

3.2.2.2. Distributed Markov Decision Processes

Participant: Serge Haddad.

Distributed systems featuring non-deterministic and probabilistic aspects are usually hard to analyze and, more specifically, to optimize. Furthermore, high complexity theoretical lower bounds have been established for models like partially observed Markovian decision processes and distributed partially observed Markovian decision processes. We believe that these negative results are consequences of the choice of the models rather than the intrinsic complexity of problems to be solved. Thus we plan to introduce new models in which the associated optimization problems can be solved in a more efficient way. More precisely, we start by studying connection protocols weighted by costs and we look for online and offline strategies for optimizing the mean cost to achieve the protocol. We have been cooperating on this subject with the SUMO team at Inria Rennes; in the joint work [32]; there, we strive to synthesize for a given MDP a control so as to guarantee a specific stationary behavior, rather than - as is usually done - so as to maximize some reward.

3.2.3. Large scale probabilistic systems

Addressing large-scale probabilistic systems requires to face state explosion, due to both the discrete part and the probabilistic part of the model. In order to deal with such systems, different approaches have been proposed:

- Restricting the synchronization between the components as in queuing networks allows to express the steady-state distribution of the model by an analytical formula called a product-form [37].

- Some methods that tackle with the combinatory explosion for discrete-event systems can be generalized to stochastic systems using an appropriate theory. For instance symmetry based methods have been generalized to stochastic systems with the help of aggregation theory [42].
- At last simulation, which works as soon as a stochastic operational semantic is defined, has been adapted to perform statistical model checking. Roughly speaking, it consists to produce a confidence interval for the probability that a random path fulfills a formula of some temporal logic [54] .

We want to contribute to these three axes: (1) we are looking for product-forms related to systems where synchronization are more involved (like in Petri nets [6]); (2) we want to adapt methods for discrete-event systems that require some theoretical developments in the stochastic framework and, (3) we plan to address some important limitations of statistical model checking like the expressiveness of the associated logic and the handling of rare events.

3.2.4. Real time distributed systems

Nowadays, software systems largely depend on complex timing constraints and usually consist of many interacting local components. Among them, railway crossings, traffic control units, mobile phones, computer servers, and many more safety-critical systems are subject to particular quality standards. It is therefore becoming increasingly important to look at networks of timed systems, which allow real-time systems to operate in a distributed manner.

Timed automata are a well-studied formalism to describe reactive systems that come with timing constraints. For modeling distributed real-time systems, networks of timed automata have been considered, where the local clocks of the processes usually evolve at the same rate [50] [41]. It is, however, not always adequate to assume that distributed components of a system obey a global time. Actually, there is generally no reason to assume that different timed systems in the networks refer to the same time or evolve at the same rate. Any component is rather determined by local influences such as temperature and workload.

3.2.4.1. Implementation of Real-Time Concurrent Systems

Participants: Thomas Chatain, Stefan Haar, Serge Haddad.

This was one of the tasks of the ANR ImpRo.

Formal models for real-time systems, like timed automata and time Petri nets, have been extensively studied and have proved their interest for the verification of real-time systems. On the other hand, the question of using these models as specifications for designing real-time systems raises some difficulties. One of those comes from the fact that the real-time constraints introduce some artifacts and because of them some syntactically correct models have a formal semantics that is clearly unrealistic. One famous situation is the case of Zeno executions, where the formal semantics allows the system to do infinitely many actions in finite time. But there are other problems, and some of them are related to the distributed nature of the system. These are the ones we address here.

One approach to implementability problems is to formalize either syntactical or behavioral requirements about what should be considered as a reasonable model, and reject other models. Another approach is to adapt the formal semantics such that only realistic behaviors are considered.

These techniques are preliminaries for dealing with the problem of implementability of models. Indeed implementing a model may be possible at the cost of some transformation, which make it suitable for the target device. By the way these transformations may be of interest for the designer who can now use high-level features in a model of a system or protocol, and rely on the transformation to make it implementable.

We aim at formalizing and automating translations that preserve both the timed semantics and the concurrent semantics. This effort is crucial for extending concurrency-oriented methods for logical time, in particular for exploiting partial order properties. In fact, validation and management - in a broad sense - of distributed systems is not realistic *in general* without understanding and control of their real-time dependent features; the link between real-time and logical-time behaviors is thus crucial for many aspects of *MExICo*'s work.

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

MExICo's research is motivated by problems of system management in several domains, such as:

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize adaptators for every component in order to steer its interaction with the surrounding processes.
- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

Currently, we have no active cooperation on these subjects.

4.2. Biological Regulation Networks

Participants: Thomas Chatain, Matthias Fuegger, Stefan Haar, Serge Haddad, Juraj Kolcak, Hugues Mandon, Stefan Schwoon.

We have begun in 2014 to examine concurrency issues in systems biology, and are currently enlarging the scope of our research's applications in this direction. To see the context, note that in recent years, a considerable shift of biologists' interest can be observed, from the mapping of static genotypes to gene expression, i.e. the processes in which genetic information is used in producing functional products. These processes are far from being uniquely determined by the gene itself, or even jointly with static properties of the environment; rather, regulation occurs throughout the expression processes, with specific mechanisms increasing or decreasing the production of various products, and thus modulating the outcome. These regulations are central in understanding cell fate (how does the cell differenciate ? Do mutations occur ? etc), and progress there hinges on our capacity to analyse, predict, monitor and control complex and variegated processes. We have applied Petri net unfolding techniques for the efficient computation of attractors in a regulatory network; that is, to identify strongly connected reachability components that correspond to stable evolutions, e.g. of a cell that differentiates into a specific functionality (or mutation). This constitutes the starting point of a broader research with Petri net unfolding techniques in regulation. In fact, the use of ordinary Petri nets for capturing regulatory network (RN) dynamics overcomes the limitations of traditional RN models : those impose e.g. Monotonicity properties in the influence that one factor had upon another, i.e. always increasing or always decreasing, and were thus unable to cover all actual behaviours. Rather, we follow the more refined model of boolean networks of automata, where the local states of the different factors jointly detemine which state transitions are possible. For these connectors, ordinary PNs constitute a first approximation, improving greatly over the literature but leaving room for improvement in terms of introducing more refined logical connectors. Future work thus involves transcending this class of PN models. Via unfoldings, one has access – provided efficient techniques are available – to all behaviours of the model, rather than over-or under-approximations as previously. This opens the way to efficiently searching in particular for determinants of the cell fate : which attractors are reachable from a given stage, and what are the factors that decide in favor of one or the other attractor, etc. Our current research focusses cellular reprogramming on the one hand, and distributed algorithms in wild or synthetic biological systems on the other. The latter is a distributed algorithms' view on microbiological systems, both with the goal to model and analyze existing microbiological systems as distributed systems, and to design and implement distributed algorithms in synthesized microbiological systems. Envisioned major long-term goals are drug production and medical treatment via synthesized bacterial colonies. We are approaching our goal of a distributed algorithm's view of microbiological systems from several directions: (i) Timing plays a crucial role in microbiological systems. Similar to modern VLSI circuits, dominating loading effects and noise render classical delay models unfeasible. In previous work we showed limitations of current delay models and presented a class of new delay models, so called involution channels. In [26] we showed that involution channels are still in accordance with Newtonian physics, even in presence of noise. (ii) In [7] we analyzed metastability in circuits by a three-valued Kleene logic, presented a general technique to build circuits that can tolerate a certain degree of metastability at its inputs, and showed the presence of a computational hierarchy. Again, we expect metastability to play a crucial role in microbiological systems, as similar to modern VLSI circuits, loading effects are pronounced. (iii) We studied agreement problems in highly dynamic networks without stability guarantees [28], [27]. We expect such networks to occur in bacterial cultures where bacteria communicate by producing and sensing small signal molecules like AHL. Both works also have theoretically relevant implications: The work in [27] presents the first approximate agreement protocol in a multidimensional space with time complexity independent of the dimension, working also in presence of Byzantine faults. In [28] we proved a tight lower bound on convergence rates and time complexity of asymptotic and approximate agreement in dynamic and classical static fault models. (iv) We are currently working with Manish Kushwaha (INRA), and Thomas Nowak (LRI) on biological infection models for E. coli colonies and M13 phages.

4.3. Metabolic Networks

Participant: Philippe Dague.

Analysis of metabolic networks in presence of biological (thermodynamical, kinetic, gene regulatory) constraints has been studied achieving a complete mathematical characterization of the solutions space at steady state (generalization of the elementary flux modes) and investigating related computing methods.

4.4. Transportation Systems

Participants: Thomas Chatain, Stefan Haar, Serge Haddad, Stefan Schwoon.

- **Autonomous Vehicles.** The validation of safety properties is a crucial concern for the design of computer guided systems, in particular for automated transport systems. Our approach consists in analyzing the interactions of a randomized environment (roads, cross-sections, etc.) with a vehicle controller.
- **Multimodal Transport Networks.** We are interested in predicting and harnessing the propagation of perturbations across different transportation modes.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- The article *Manifestability Verification of Discrete Event Systems* by Lina Ye, Philippe Dague, and Lulu He received the *Best Paper Award* of the *30th International Workshop on Principles of Diagnosis DX'19*, Klagenfurt/Austria, November 2019.
- The article *Sequential Reprogramming of Boolean Networks Made Practical* by Hugues Mandon, Cui Su, Stefan Haar, Jun Pang, and Loïc Paulevé received the *Best Paper Award* of the conference on *Computational Models in Systems Biology*(CMSB 2019), Trieste/Italy, September 18-20, 2019.

BEST PAPERS AWARDS :

[24]
L. YE, P. DAGUE, L. HE.*Manifestability Verification of Discrete Event Systems*, in "DX'19 - 30th International Workshop on Principles of Diagnosis", Klagenfurt, Austria, November 2019, vol. 19, p. 1-9, Best Paper Award (https://dx-workshop.org/2019/awards/), https://hal.archives-ouvertes.fr/hal-02425146

[22]

H. MANDON, C. SU, S. HAAR, J. PANG, L. PAULEVÉ.*Sequential Reprogramming of Boolean Networks Made Practical*, in "CMSB 2019 - 17th International Conference on Computational Methods in Systems Biology", Trieste, France, Lecture Notes in Computer Science, Springer, 2019, vol. 11773, p. 3–19, Best paper award [*DOI :* 10.1007/978-3-030-31304-3_1], https://hal.archives-ouvertes.fr/hal-02178917

6. New Software and Platforms

6.1. COSMOS

KEYWORD: Model Checker

FUNCTIONAL DESCRIPTION: COSMOS is a statistical model checker for the Hybrid Automata Stochastic Logic (HASL). HASL employs Linear Hybrid Automata (LHA), a generalization of Deterministic Timed Automata (DTA), to describe accepting execution paths of a Discrete Event Stochastic Process (DESP), a class of stochastic models which includes, but is not limited to, Markov chains. As a result HASL verification turns out to be a unifying framework where sophisticated temporal reasoning is naturally blended with elaborate reward-based analysis. COSMOS takes as input a DESP (described in terms of a Generalized Stochastic Petri Net), an LHA and an expression Z representing the quantity to be estimated. It returns a confidence interval estimation of Z, recently, it has been equipped with functionalities for rare event analysis.

It is easy to generate and use a C code for discrete Simulink models (using only discrete blocks, which are sampled at fixed intervals) using MathWorks tools. However, it limits the expressivity of the models. In order to use more diverse Simulink models and control the flow of a multi-model simulation (with Discrete Event Stochastic Processes) we developed a Simulink Simulation Engine embedded into Cosmos.

COSMOS is written in C++

- Participants: Benoît Barbot, Hilal Djafri, Marie Duflot-Kremer, Paolo Ballarini and Serge Haddad
- Contact: Benoît Barbot
- URL: http://www.lsv.ens-cachan.fr/~barbot/cosmos/

6.2. CosyVerif

FUNCTIONAL DESCRIPTION: CosyVerif is a platform dedicated to the formal specification and verification of dynamic systems. It allows to specify systems using several formalisms (such as automata and Petri nets), and to run verification tools on these models.

- Participants: Alban Linard, Fabrice Kordon, Laure Petrucci and Serge Haddad
- Partners: LIP6 - LSV - LIPN (Laboratoire d'Informatique de l'Université Paris Nord)
- Contact: Serge Haddad
- URL: http://www.cosyverif.org/

6.3. Mole

FUNCTIONAL DESCRIPTION: Mole computes, given a safe Petri net, a finite prefix of its unfolding. It is designed to be compatible with other tools, such as PEP and the Model-Checking Kit, which are using the resulting unfolding for reachability checking and other analyses. The tool Mole arose out of earlier work on Petri nets.

- Participant: Stefan Schwoon
- Contact: Stefan Schwoon
- URL: http://www.lsv.ens-cachan.fr/~schwoon/tools/mole/

7. New Results

7.1. Generalized Alignment-Based Trace Clustering of Process Behavior

Process mining techniques use event logs containing real process executions in order to mine, align and extend process models. The partition of an event log into trace variants facilitates the understanding and analysis of traces, so it is a common pre-processing in process mining environments. Trace clustering automates this partition; traditionally it has been applied without taking into consideration the availability of a process model. In this paper we extend our previous work on process model based trace clustering, by allowing cluster centroids to have a complex structure, that can range from a partial order, down to a subnet of the initial process model. This way, the new clustering framework presented in [28] is able to cluster together traces that are distant only due to concurrency or loop constructs in process models. We show the complexity analysis of the different instantiations of the trace clustering framework, and have implemented it in a prototype tool that has been tested on different datasets.

7.2. The involution tool for accurate digital timing and power analysis

In [23] we introduce the prototype of a digital timing simulation and power analysis tool for integrated circuit (Involution Tool) which employs the involution delay model introduced by Fuegger et al. at DATE'15. Unlike the pure and inertial delay models typically used in digital timing analysis tools, the involution model faithfully captures pulse propagation. The presented tool is able to quantify for the first time the accuracy of the latter by facilitating comparisons of its timing and power predictions with both SPICE-generated results and results achieved by standard timing analysis tools. It is easily customizable, both w.r.t. different instances of the involution model and different circuits, and supports automatic test case generation, including parameter sweeping. We demonstrate its capabilities by providing timing and power analysis results for three circuits in varying technologies: an inverter tree, the clock tree of an open-source processor, and a combinational circuit that involves multi-input NAND gates. It turns out that the timing and power predictions of two natural types of involution models are significantly better than the predictions obtained by standard digital simulations for the inverter tree and the clock tree. For the NAND circuit, the performance is comparable but not significantly better. Our simulations thus confirm the benefits of the involution model, but also demonstrate shortcomings for multi-input gates.

7.3. Transistor-level analysis of dynamic delay models

Delay estimation is a crucial task in digital circuit design as it provides the possibility to assure the desired func-tionality, but also prevents undesired behavior very early. For this purpose elaborate delay models like the Degradation Delay Model (DDM) and the Involution Delay Model (IDM) have been proposed in the past, which facilitate accurate dynamic timing analysis: Both use delay functions that determine the delay of the current input transition based on the time difference T to the previous output one. Currently, however, extensive analog simulations are necessary to determine the (parameters of the) delay function, which is a very time-consuming and cumbersome task and thus limits the applicability of these models. In [21], we therefore thoroughly investigate the characterization procedures of a CMOS inverter on the transistor level in order to derive analytical expressions for the delay functions. Based on reasonably simple transistor models we identify three operation regions, each described by a different estimation function. Using simulations with two independent technologies, we show that our predictions are not only accurate but also reasonably robust w.r.t. variations. Our results furthermore indicate that the exponential fitting proposed for DDM is actually only partially valid, while our analytic approach can be applied on the whole range. Even the more complex IDM is predicted reasonably accurate.

7.4. A faithful binary circuit model

[Fuegger et al., IEEE TC 2016] proved that no existing digital circuit model, including those based on pure and inertial delay channels, faithfully captures glitch propagation: For the Short-Pulse Filtration (SPF) problem similar to that of building a one-shot inertial delay, they showed that every member of the broad class of bounded single-history channels either contradicts the unsolvability of SPF in bounded time or the solvability of SPF in unbounded time in physical circuits. In [12], we propose binary circuit models based on novel involution channels that do not suffer from this deficiency. Namely, in sharp contrast to bounded single-history channels, SPF cannot be solved in bounded time with involution channels, whereas it is easy to provide an unbounded SPF implementation. Hence, binary-valued circuit models based on involution channels allow to solve SPF precisely when this is possible in physical circuits. Additionally, using both Spice simulations and physical measurements of an inverter chain instrumented by high-speed analog amplifiers, we demonstrate that our model provides good modeling accuracy with respect to real circuits as well. Consequently , our involution channel model is not only a promising basis for sound formal verification, but also allows to seamlessly improve existing dynamic timing analysis.

7.5. Concurrency in Boolean networks

Boolean networks (BNs) are widely used to model the qualitative dynamics of biological systems. Besides the logical rules determining the evolution of each component with respect to the state of its regulators, the scheduling of component updates can have a dramatic impact on the predicted behaviours. In [10], we explore the use of Read (contextual) Petri Nets (RPNs) to study dynamics of BNs from a concurrency theory perspective. After showing bi-directional translations between RPNs and BNs and analogies between results on synchronism sensitivity, we illustrate that usual updating modes for BNs can miss plausible behaviours, i.e., incorrectly conclude on the absence/impossibility of reaching specific configurations. We propose an encoding of BNs capitalizing on the RPN semantics enabling more behaviour than the generalized asynchronous updating mode. The proposed encoding ensures a correct abstraction of any multivalued refinement, as one may expect to achieve when modelling biological systems with no assumption on its time features.

7.6. Sequential Reprogramming of Boolean Networks Made Practical

We address the sequential reprogramming of gene regulatory networks modelled as Boolean networks.

- Cellular reprogramming, a technique that opens huge opportunities in modern and regenerative medicine, heavily relies on identifying key genes to perturb. Most of the existing computational methods for controlling which attractor (steady state) the cell will reach focus on finding mutations to apply to the initial state. However, it has been shown, and is proved in our article [14], that waiting between perturbations so that the update dynamics of the system prepares the ground, allows for new reprogramming strategies. To identify such sequential perturbations, we consider a qualitative model of regulatory networks, and rely on Binary Decision Diagrams to model their dynamics and the putative perturbations. Our method establishes a set identification of sequential perturbations, whether permanent (mutations) or only temporary, to achieve the existential or inevitable reachability of an arbitrary state of the system. We apply an implementation for temporary perturbations on models from the literature, illustrating that we are able to derive sequential perturbations to achieve trans-differentiation.
- In [22], we develop an attractor-based sequential reprogramming method to compute all sequential reprogramming paths from a source attractor to a target attractor, where only attractors of the network are used as intermediates. Our method is more practical than existing reprogramming methods as it incorporates several practical constraints: (1) only biologically observable states, viz. attractors, can act as intermediates; (2) certain attractors, such as apoptosis, can be avoided as intermediates; (3) certain nodes can be avoided to perturb as they may be essential for cell survival or difficult to perturb with biomolecular techniques; and (4) given a threshold k, all sequential reprogramming paths with no more than k perturbations are computed. We compare our method with the minimal one-step

reprogramming and the minimal sequential reprogramming on a variety of biological networks. The results show that our method can greatly reduce the number of perturbations compared to the one-step reprogramming, while having comparable results with the minimal sequential reprogramming. Moreover, our implementation is scalable for networks of more than 60 nodes.

7.7. Parameter Space Abstraction and Unfolding Semantics of Discrete Regulatory Networks.

The modelling of discrete regulatory networks combines a graph specifying the pairwise influences between the variables of the system, and a parametrisation from which can be derived a discrete transition system. Given the influence graph only, the exploration of admissible parametrisations and the behaviours they enable is computationally demanding due to the combinatorial explosions of both parametrisation and reachable state space. In [13], we introduce an abstraction of the parametrisation space and its refinement to account for the existence of given transitions, and for constraints on the sign and observability of influences. The abstraction uses a convex sub-lattice containing the concrete parametrisation space specified by its infimum and supremum parametrisations. It is shown that the computed abstractions are optimal, i.e., no smaller convex sublattice exists. Although the abstraction may introduce over-approximation, it has been proven to be conservative with respect to reachability of states. Then, an unfolding semantics for Parametric Regulatory Networks is defined, taking advantage of concurrency between transitions to provide a compact representation of reachable transitions. A prototype implementation is provided: it has been applied to several examples of Boolean and multi-valued networks, showing its tractability for networks with numerous components.

7.8. Combining Refinement of Parametric Models with Goal-Oriented Reduction of Dynamics

Parametric models abstract part of the specification of dynamical models by integral parameters. They are for example used in computational systems biology, notably with parametric regulatory networks, which specify the global architecture (interactions) of the networks, while parameterising the precise rules for drawing the possible temporal evolutions of the states of the components. A key challenge is then to identify the discrete parameters corresponding to concrete models with desired dynamical properties. Our work [20] addresses the restriction of the abstract execution of parametric regulatory (discrete) networks by the means of static analysis of reachability properties (goal states). Initially defined at the level of concrete parameterised models, the goal-oriented reduction of dynamics is lifted to parametric networks, and is proven to preserve all the minimal traces to the specified goal states. It results that one can jointly perform the refinement of parametric networks (restriction of domain of parameters) while reducing the necessary transitions to explore and preserving reachability properties of interest.

7.9. Autonomous Transitions Enhance CSLTA Expressiveness and Conciseness

CSLTA is a stochastic temporal logic for continuous-time Markov chains (CTMC) where formulas similarly to those of CTL* are inductively defined by nesting of timed path formulas and state formulas. In particular a timed path formula of CSLTA is specified by a single-clock Deterministic Timed Automaton (DTA). Such a DTA features two kinds of transitions: synchronizing transitions triggered by CTMC transitions and autonomous transitions triggered by time elapsing that change the location of the DTA when the clock reaches a given threshold. It has already been shown that CSLTA strictly includes stochastic logics like CSL and asCSL. An interesting variant of CSLTA consists in equipping transitions rather than locations by boolean formulas. In [27], we answer the following question: do autonomous transitions and/or boolean guards on transitions enhance expressiveness and/or conciseness of DTAs? We show that this is indeed the case. In establishing our main results we also identify an accurate syntactical characterization of DTAs for which the autonomous transitions do not add expressive power but lead to exponentially more concise DTAs.

7.10. Coverability and Termination in Recursive Petri Nets

In the early two-thousands, Recursive Petri nets have been introduced in order to model distributed planning of multi-agent systems for which counters and recursivity were necessary. Although Recursive Petri nets strictly extend Petri nets and stack automata, most of the usual property problems are solvable but using non primitive recursive algorithms, even for coverability and termination. For almost all other extended Petri nets models containing a stack the complexity of coverability and termination are unknown or strictly larger than EXPSPACE. In contrast, we establish in [18] that for Recursive Petri nets, the coverability and termination problems are EXPSPACE-complete as for Petri nets. From an expressiveness point of view, we show that coverability languages of Recursive Petri nets strictly include the union of coverability languages of Petri nets and context-free languages. Thus we get for free a more powerful model than Petri net.

8. Partnerships and Cooperations

8.1. Regional Initiatives

- MATTHIAS FUEGGER is co-leading the Digicosme working group *HicDiesMeus* on *Highly Constrained Discrete Agents for Modeling Natural Systems*.
- STEFAN HAAR is co-leading the Digicosme working group *TheoBioR* on *Computational methods for modelling and analysing biological networks*.

8.2. National Initiatives

- Thomas Chatain, Stefan Haar, Serge Haddad and Stefan Schwoon are participating in the ANR Project ALGORECELL.
- Matthias Függer participates in the ANR project FREDDA on verification and synthesis of distributed algorithms.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

- Susanna DONATELLI was invited professor of ENS Paris-Saclay during one month in January, working with Serge Haddad on the expressiveness and conciseness of temporal logic for Markov chains. This work was also coninued durin a visit of Serge Haddad at the university of Torino in March. Their joint work has led to a publication to appear in the international conference LATA 2020 at Milano.
- Sven DZIADEK, Sep-Nov 2019 (PhD student, Univ. Leipzig)

8.3.1.1. Research Stays Abroad

- JURAJ KOLČÁK visited the SDM group of Hasuo Ichiro at NII Tokyo from August 2018 to February 2019, working in particular on differential logics.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. General Chair, Scientific Chair

- Serge Haddad is a member of the stering committee of the International Conference on Application and Theory of Petri Nets and Concurrency.

- PHILIPPE DAGUE co-organized with Franck Delaplace the conference on *Computational Systems Biology for Complex Diseases* on November 28-29, 2019 at ENS Paris-Saclay, which gathered around 80 participants.
- PHILIPPE DAGUE, as responsible of the working group BIOSS-IA (Systemic Symbolic Biology and Artificial Intelligence) of the GDR IA, organized the 2019 Day at Orléans on May 27.
- MATTHIAS FUEGGER co-organized the CELLS workshop at DISC'19. The workshop covers topics from computing among a consortium of cells.
- STEFAN HAAR co-directed with Benedikt Bollig the scientific and logistic organisation of the ForMal workshop at ENS Paris-Saclay on cross-fertilization between formal methods and machine learning.

9.1.1.2. Member of the Organizing Committees

- STEFAN HAAR was co-chair and SERGE HADDAD was a member of the Scientific Committee (including organisational issues) of the Digicosme Spring School on Formal Methods and Machine Learning held in June at Cachan *(ForMaL)*.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- STEFAN HAAR was co-chair of program committee for the *40th International Conference on Application and Theory of Petri Nets and Concurrency*, Aachen, Germany, June 23-28, 2019.

9.1.2.2. Member of the Conference Program Committees

- Serge Haddad was a member of the PC of
  - the workshop PNSE associated with ATPN 2019, Aachen, Germany, and
  - the 12th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS 2019), Palma de Mallorca, Spain.
- PHILIPPE DAGUE was a member of the program committee of the *30th International Workshop on Principles of Diagnosis DX'19*, Klagenfurt/Austria, November 2019.
- STEFAN HAAR was
  - a member of the programm committee of the *19th International Conference on Applications of Concurrency to Systems Design (ACSD 2019)*, Aachen, Germany, June 23-28, 2019, and
  - a member of the program committee of the *Workshop Algorithms & Theories for the Analysis of Event Data 2019* (ATAED 2019).
- MATTHIAS FUEGGER was
  - Steering commitee member of IEEE ASYNC'19 and
  - PC member of IEEE ASYNC'19 and IEEE DDECS'19.
- THOMAS CHATAIN was
  - a member of the programm committee of the *19th International Conference on Applications of Concurrency to Systems Design (ACSD 2019)*, Aachen, Germany, June 23-28, 2019, and
  - a member of the program committee of the *1st International Conference on Process Mining (ICPM 2019)*
- LINA YE was a member of the program committee of the *17th International Workshop on Coordination and Self-adaptativeness of Software Applications* (FOCLASA 2019).
- Stefan Schwoon acted as reviewer for MFCS 2019, FSTTCS 2019, FOSSACS 2019,, TACAS 2019, and STACS 2020.
- LINA YE was

- a reviewer of the *58th Conference on Decision and Control* (CDC 2019), and
- a reviewer of the *the 30th International Workshop on Principles of Diagnosis* (DX 2019).

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- STEFAN HAAR is an associate editor for *Journal of Discrete Events Dynamic Systems: Theory and Application* (JDEDS).

9.1.3.2. Reviewer - Reviewing Activities

- Serge Haddad was a reviewer for the following journals:
  - Journal of Logical and Algebraic Methods in Programming,
  - Innovations in Systems and Software Engineering, and
  - Transactions on Petri Nets and Other Models of Concurrency.
- Thomas Chatain was a reviewer for journals *Journal of Discrete Events Dynamic Systems: Theory and Application* (JDEDS) and Fundamenta Informaticae.
- Stefan Schwoon was a reviewer for *Journal of Discrete Event Dynamic Systems* (JDEDS), *Transactions on Software Engineering* (TSE), *Petri Nets and Other Models of Concurrency* (ToPNoC), and Transactions on Programming Languages and Systems (TOPLAS).
- LINA YE is a reviewer for *Journal of IEEE Access* (IEEE Access).

9.1.4. Invited Talks

- Stefan Schwoon gave a talk in the SYBILA seminar of Masaryk University, Brno, on *Diagnosis and Opacity in Partially Observable Systems*.

9.1.5. Scientific Expertise

- Serge Haddad is a member of the scientific and administrative council (CSA) of Labex CIMI of Toulouse and a member of the scientific orientation council (COS) of LIS of Marseille (UMR 7020).

9.1.6. Research Administration

- Serge Haddad is Head of the Computer Science Department of ENS Paris-Saclay.
- Stefan Haar is the president of Inria's COST-GTRI.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- MATHIAS FUEGGER,

  Master :

  - *Initiation à la recherche*, 10 h EQTD, M1, ENS Paris-Saclay, France
- STEFAN HAAR,

  Licence : *Langages Formels*, EQTD, L3, ENS Paris-Saclay, France;

  Master :

  - *Analyse de la dynamique des systèmes biologiques*, 10 h EQTD, M1, Université PAris-Saclay, France
  - *Initiation à la recherche*, 10 h EQTD, M1, ENS Paris-Saclay, France
  - *Module SPECIF* , 12 h EQTD, M1, UPMC, France.

  Serge Haddad is head of the Computer Science department of ENS Paris-Saclay. He teaches basic and advanced algorithmics (L3) and probabilistic features of computer science (M1).

STEFAN SCHWOON

– Responsable L3 Informatique, ENS Paris-Saclay
– Enseignement au M1 MPRI : cours *Initiation à la Vérification* (22,5h)
– Enseignement au L3 Info : cours *Architecture et Système* (45h), projet *Programmation orienté objet* (15h), TD *Langages Formels* (22,5h)
– Enseignement à l'Aggrétation Maths Option Informatique: cours *Algorithmique* (22,5h)

9.2.2. Supervision

- SERGE HADDAD is supervising with Alain Finkel the PhD thesis of Igor Khmelnitsky on Verification of infinite-state systems and machine learning.
- STEFAN HAAR has been supervising, with Co-superviser Loic Paulevé at LABRI, the PhD theses of

  HUGUES MANDON, *Models and Algorithms for cellular reprogramming strategies prediction*, ENS Paris-Saclay, defended on Nov. 19, 2019, and

  JURAJ KOLCˇÁK, *Parametric Logical Regulatory Networks*, PhD research started in March 2017.
- THOMAS CHATAIN has been supervising, with co-superviser Josep Carmona at Universitat Politècnica de Catalunya (Barcelona, Spain), the PhD thesis of MATHILDE BOLTENHAGEN, *Optimization Techniques for Conformance Checking and Model Repair in Process Mining*, PhD research started in November 2018.
- LINA YE has been supervising, with Co-superviser Philippe Dague at LRI, the PhD these of LULU HE, *Robustness Analysis of Real-Time Systems*, PhD research started in February 2019.

9.2.3. Juries

- PHILIPPE DAGUE was *garant* and member of the HdR jury of Sabine Peres, and a member of the HdR jury of Jean-Marie Lagniez.
- Serge Haddad was the president of the PhD committee of Mauricio Gonzalez defended at ENS Paris-Saclay in November.
- THOMAS CHATAIN has been a jury member for the PhD defense of SAMY JAZIRI (supervisers Patricia Bouyer-Decitre et Nicolas Markey) at Université Paris-Saclay in September 2019.
- THOMAS CHATAIN has been a jury member for the SIF PhD award (Prix de thèse Gilles Kahn 2019).

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

- STEFAN HAAR has been the president of COST-GTRI.

10. Bibliography

Major publications by the team in recent years

[1] É. ANDRÉ, TH. CHATAIN, C. RODRÍGUEZ.*Preserving Partial-Order Runs in Parametric Time Petri Nets*, in "ACM Transactions in Embedded Computing Systems", 2017, vol. 16, n^{o} 2, p. 1–26

[2] B. BÉRARD, S. HAAR, S. SCHMITZ, S. SCHWOON.*The Complexity of Diagnosability and Opacity Verification for Petri Nets*, in "Fundamenta Informaticae", 2018, vol. 161, n^{o} 4, p. 317-349 [*DOI :* 10.3233/FI-2018-1706]

[3] TH. CHATAIN, S. HAAR, J. KOLČÁK, L. PAULEVÉ, A. THAKKAR.*Concurrency in Boolean networks*, in "Natural Computing", 2019

[4] S. FRIEDRICHS, M. FÜGGER, C. LENZEN.*Metastability-Containing Circuits*, in "IEEE Transactions on Computers", 2018, vol. 67, n^{o} 8 [*DOI :* 10.1109/TC.2018.2808185]

[5] S. HAAR, S. HADDAD, T. MELLITI, S. SCHWOON.*Optimal constructions for active diagnosis*, in "Journal of Computer and System Sciences", 2017, vol. 83, n^{o} 1, p. 101-120

[6] S. HADDAD, J. MAIRESSE, H.-T. NGUYEN.*Synthesis and Analysis of Product-form Petri Nets*, in "Fundamenta Informaticae", 2013, vol. 122, n^{o} 1-2, p. 147-172

[7] J. KOLČÁK, D. ŠAFRÁNEK, S. HAAR, L. PAULEVÉ.*Parameter Space Abstraction and Unfolding Semantics of Discrete Regulatory Networks*, in "Theoretical Computer Science", 2019, vol. 765, p. 120-144, https://hal.archives-ouvertes.fr/hal-01734805

Publications of the year

Doctoral Dissertations and Habilitation Theses

[8] H. MANDON.*Algorithms for Cell Reprogramming Strategies in Boolean Networks*, Université Paris-Saclay, November 2019, https://hal.archives-ouvertes.fr/tel-02412717

Articles in International Peer-Reviewed Journal

[9] N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*A Tale of Two Diagnoses in Probabilistic Systems*, in "Journal of Information and Computation", December 2019, vol. 269, p. 1-33 [*DOI :* 10.1016/J.IC.2019.104441], https://hal.inria.fr/hal-02430814

[10] T. CHATAIN, S. HAAR, J. KOLČÁK, L. PAULEVÉ, A. THAKKAR.*Concurrency in Boolean networks*, in "Natural Computing", 2019 [*DOI :* 10.1007/S11047-019-09748-4], https://hal.inria.fr/hal-01893106

[11] P. DAGUE, L. HE, L. YE.*How to be sure a faulty system does not always appear healthy?Fault manifestability analysis for discrete event and timed systems*, in "Innovations in Systems and Software Engineering", 2019, forthcoming [*DOI :* 10.1007/S11334-019-00357-Z], https://hal.archives-ouvertes.fr/hal-02425142

[12] M. FÜGGER, R. NAJVIRT, T. NOWAK, U. SCHMID.*A Faithful Binary Circuit Model*, in "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems", August 2019 [*DOI :* 10.1109/TCAD.2019.2937748], https://hal.inria.fr/hal-02395251

[13] J. KOLČÁK, D. ŠAFRÁNEK, S. HAAR, L. PAULEVÉ.*Parameter Space Abstraction and Unfolding Semantics of Discrete Regulatory Networks*, in "Theoretical Computer Science", December 2019, vol. 765, p. 120-144 [*DOI :* 10.1016/J.TCS.2018.03.009], https://hal.archives-ouvertes.fr/hal-01734805

[14] H. MANDON, C. SU, J. PANG, S. PAUL, S. HAAR, L. PAULEVÉ.*Algorithms for the Sequential Reprogramming of Boolean Networks*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", 2019, vol. 16, n^{o} 5, p. 1610-1619 [*DOI :* 10.1109/TCBB.2019.2914383], https://hal.archives-ouvertes.fr/hal-02113864

[15] W. M. VAN DER AALST, J. CARMONA, T. CHATAIN, B. F. VAN DONGEN.*A Tour in Process Mining: From Practice to Algorithmic Challenges*, in "LNCS Transactions on Petri Nets and Other Models of Concurrency", December 2019 [*DOI :* 10.1007/978-3-662-60651-3_1], https://hal.inria.fr/hal-02419141

International Conferences with Proceedings

[16] M. BOLTENHAGEN, T. CHATAIN, J. CARMONA.*Encoding Conformance Checking Artefacts in SAT*, in "15th International Workshop on Business Process Intelligence (BPI'19)", Wien, Austria, September 2019, https://hal.inria.fr/hal-02419980

[17] M. BOLTENHAGEN, T. CHATAIN, J. CARMONA.*Generalized Alignment-Based Trace Clustering of Process Behavior*, in "Petri Nets 2019 / ACSD 2019 - 40th International Conference on Application and Theory of Petri Nets and Concurrency", Aachen, Germany, Application and Theory of Petri Nets and Concurrency 40th International Conference, PETRI NETS 2019, Aachen, Germany, June 23–28, 2019, Proceedings, June 2019, https://hal.archives-ouvertes.fr/hal-02176771

[18] A. FINKEL, S. HADDAD, I. KHMELNITSKY.*Coverability and Termination in Recursive Petri Nets*, in "Petri Nets 2019 / ACSD 2019 - 40th International Conference on Application and Theory of Petri Nets and Concurrency", Aachen, Germany, June 2019, https://hal.inria.fr/hal-02081019

[19] A. FINKEL, S. HADDAD, I. KHMELNITSKY.*Réification des accélérations pour la construction de Karp et Miller*, in "MSR 2019 - 12ème Colloque sur la Modélisation des Systèmes Réactifs", Angers, France, November 2019, https://hal.inria.fr/hal-02319755

[20] S. HAAR, J. KOLČÁK, L. PAULEVÉ.*Combining Refinement of Parametric Models with Goal-Oriented Reduction of Dynamics*, in "VMCAI 2019 - 20th International Conference on Verification, Model Checking, and Abstract Interpretation", Lisbon, Portugal, Lecture Notes in Computer Science, Springer, 2019, vol. 11388, p. 555-576, https://arxiv.org/abs/1811.12377 [*DOI :* 10.1007/978-3-030-11245-5_26], https://hal.archives-ouvertes.fr/hal-01940174

[21] J. MAIER, M. FÜGGER, T. NOWAK, U. SCHMID.*Transistor-Level Analysis of Dynamic Delay Models*, in "ASYNC 2019 - 25th IEEE International Symposium on Asynchronous Circuits and Systems", Hirosaki, Japan, IEEE, May 2019, p. 76-85 [*DOI :* 10.1109/ASYNC.2019.00019], https://hal.inria.fr/hal-02395229

[22] *Best Paper*
H. MANDON, C. SU, S. HAAR, J. PANG, L. PAULEVÉ.*Sequential Reprogramming of Boolean Networks Made Practical*, in "CMSB 2019 - 17th International Conference on Computational Methods in Systems Biology", Trieste, France, Lecture Notes in Computer Science, Springer, 2019, vol. 11773, p. 3–19, Best paper award [*DOI :* 10.1007/978-3-030-31304-3_1], https://hal.archives-ouvertes.fr/hal-02178917.

[23] D. OHLINGER, J. MAIER, M. FÜGGER, U. SCHMID.*The Involution Tool for Accurate Digital Timingand Power Analysis*, in "PATMOS 2019 - 29th International Symposium on Power and Timing Modeling, Optimization and Simulation", Rhodes, Greece, IEEE, July 2019 [*DOI :* 10.1109/PATMOS.2019.8862165], https://hal.inria.fr/hal-02395242

[24] *Best Paper*
L. YE, P. DAGUE, L. HE.*Manifestability Verification of Discrete Event Systems*, in "DX'19 - 30th International Workshop on Principles of Diagnosis", Klagenfurt, Austria, November 2019, vol. 19, p. 1-9, Best Paper Award (https://dx-workshop.org/2019/awards/), https://hal.archives-ouvertes.fr/hal-02425146.

[25] D. ZONETTI, A. SAOUD, A. GIRARD, L. FRIBOURG.*A symbolic approach to voltage stability and power sharing in time-varying DC microgrids*, in "ECC 2019 - European control conference", Naples, Italy, June 2019, European Control Conference 2019 [*DOI :* 10.23919/ECC.2019.8796095], https://hal.archives-ouvertes.fr/hal-02070070

[26] D. ZONETTI, A. SAOUD, A. GIRARD, L. FRIBOURG.*Decentralized monotonicity-based voltage control of DC microgrids with ZIP loads*, in "NecSys 2019 - 8th IFAC Workshop on Distributed Estimation and Control in Networked Systems", Chicago, United States, September 2019, 8th IFAC Workshop on Distributed Estimation and Control in Networked Systems (NECSYS) 2019, https://hal.archives-ouvertes.fr/hal-02208665

Research Reports

[27] S. DONATELLI, S. HADDAD.*Autonomous Transitions Enhance CSLTA Expressiveness and Conciseness*, Inria Saclay Ile de France ; LSV, ENS Cachan, CNRS, Inria, Université Paris-Saclay, Cachan (France) ; Universita degli Studi di Torino, October 2019, https://hal.inria.fr/hal-02306021

Other Publications

[28] T. CHATAIN, M. BOLTENHAGEN, J. CARMONA.*Anti-Alignments – Measuring The Precision of Process Models and Event Logs*, November 2019, https://arxiv.org/abs/1912.05907 - working paper or preprint, https://hal.inria.fr/hal-02383546

[29] A. SAOUD, A. GIRARD, L. FRIBOURG.*Assume-guarantee contracts for discrete and continuous-time systems*, July 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02196511

[30] A. SAOUD, A. GIRARD, L. FRIBOURG.*Contract-based Design of Symbolic Controllers for Safety in Distributed Multiperiodic Sampled-Data Systems*, May 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02132070

References in notes

[31] S. ABBES, A. BENVENISTE, S. HAAR.*A Petri net model for distributed estimation*, in "Proc. MTNS 2004, Sixteenth International Symposium on Mathematical Theory of Networks and Systems, Louvain (Belgium), ISBN 90-5682-517-8", 2004

[32] S. AKSHAY, N. BERTRAND, S. HADDAD, L. HELOUET.*The steady-state control problem for Markov decision processes*, in "Qest 2013", Buenos Aires, Argentina, K. R. JOSHI, M. SIEGLE, M. STOELINGA, P. R. D'ARGENIO (editors), Springer, September 2013, vol. 8054, p. 290-304, https://hal.inria.fr/hal-00879355

[33] R. ALUR, K. ETESSAMI, M. YANNAKAKIS.*Realizability and Verification of MSC Graphs*, in "Theor. Comput. Sci.", 2005, vol. 331, n^{o} 1, p. 97–114

[34] P. BALDAN, TH. CHATAIN, S. HAAR, B. KÖNIG.*Unfolding-based Diagnosis of Systems with an Evolving Topology*, in "Information and Computation", October 2010, vol. 208, n^{o} 10, p. 1169-1192, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BCHK-icomp10.pdf

[35] P. BALDAN, A. CORRADINI, B. KÖNIG, S. SCHWOON.*McMillan's complete prefix for contextual nets*, in "Transactions on Petri Nets and Other Models of Concurrency", November 2008, vol. 1, p. 199–220, Volume 5100 of Lecture Notes in Computer Science

[36] P. BALDAN, S. HAAR, B. KOENIG.*Distributed Unfolding of Petri Nets*, in "Proc.FOSSACS 2006", LNCS, Springer, 2006, vol. 3921, p. 126-141, Extended version: Technical Report CS-2006-1. Department of Computer Science, University Ca' Foscari of Venice

[37] F. BASKETT, K. M. CHANDY, R. R. MUNTZ, F. G. PALACIOS.*Open, Closed, and Mixed Networks of Queues with Different Classes of Customers*, in "J. ACM", April 1975, vol. 22, p. 248–260, http://doi.acm.org/10.1145/321879.321887

[38] A. BENVENISTE, É. FABRE, S. HAAR.*Markov Nets: Probabilistic Models for distributed and concurrent Systems*, in "IEEE Transactions on Automatic Control", 2003, vol. 48 (11), p. 1936-1950, Extended version: IRISA Research Report 1538

[39] P. BHATEJA, P. GASTIN, M. MUKUND, K. NARAYAN KUMAR.*Local testing of message sequence charts is difficult*, in "Proceedings of the 16th International Symposium on Fundamentals of Computation Theory (FCT'07)", Budapest, Hungary, E. CSUHAJ-VARJÚ, Z. ÉSIK (editors), Lecture Notes in Computer Science, Springer, August 2007, vol. 4639, p. 76-87 [*DOI :* 10.1007/978-3-540-74240-1_8], http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BGMN-fct07.pdf

[40] A. BOUILLARD, S. HAAR, S. ROSARIO.*Critical paths in the Partial Order Unfolding of a Stochastic Petri Net*, in "Proceedings of the 7th International Conference on Formal Modelling and Analysis of Timed Systems (FORMATS'09)", Budapest, Hungary, J. OUAKNINE, F. VAANDRAGER (editors), Lecture Notes in Computer Science, Springer, September 2009, vol. 5813, p. 43-57 [*DOI :* 10.1007/978-3-642-04368-0_6], http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BHR-formats09.pdf

[41] P. BOUYER, S. HADDAD, P.-A. REYNIER.*Timed Unfoldings for Networks of Timed Automata*, in "Proceedings of the 4th International Symposium on Automated Technology for Verification and Analysis (ATVA'06)", Beijing, ROC, S. GRAF, W. ZHANG (editors), Lecture Notes in Computer Science, Springer, October 2006, vol. 4218, p. 292-306, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BHR-atva06.pdf

[42] G. CHIOLA, C. DUTHEILLET, G. FRANCESCHINIS, S. HADDAD.*Stochastic Well-Formed Colored Nets and Symmetric Modeling Applications*, in "IEEE Transactions on Computers", November 1993, vol. 42, n^{o} 11, p. 1343-1360, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PS/CDFH-toc93.ps

[43] R. DEBOUK, D. TENEKETZIS.*Coordinated decentralized protocols for failure diagnosis of discrete-event systems*, in "Journal of Discrete Event Dynamical Systems: Theory and Application", 2000, vol. 10, p. 33–86

[44] J. ESPARZA, K. HELJANKO.*Unfoldings - A Partial-Order Approach to Model Checking*, EATCS Monographs in Theoretical Computer Science, Springer, 2008

[45] É. FABRE, A. BENVENISTE, C. JARD, S. HAAR.*Diagnosis of Asynchronous Discrete Event Systems, a Net Unfolding Approach*, in "IEEE Trans. Aut. Control", 2003, vol. 48 (5), p. 714-727

[46] É. FABRE, A. BENVENISTE, C. JARD, S. HAAR.*Distributed monitoring of concurrent and asynchronous systems*, in "Discrete Event Dynamic Systems: theory and application", 2005, vol. 15 (1), p. 33-84, Preliminary version: Proc. CONCUR 2003, LNCS 2761, pp.1–28, Springer

[47] S. HAAR, A. BENVENISTE, É. FABRE, C. JARD.*Partial Order Diagnosability Of Discrete Event Systems Using Petri Net Unfoldings*, in "42nd IEEE Conference on Decision and Control (CDC)", 2003

[48] S. HAAR.*Probabilistic Cluster Unfoldings*, in "Fundamenta Informaticae", 2003, vol. 53 (3-4), p. 281-314

[49] S. LAFORTUNE, Y. WANG, T.-S. YOO.*Diagnostic Decentralisé Des Systèmes A Evénements Discrets*, in "Journal Europeen des Systèmes Automatisés (RS-JESA)", August 2005, vol. 99, n° 99, p. 95–110

[50] K. G. LARSEN, P. PETTERSSON, W. YI.*Compositional and symbolic model-checking of real-time systems*, in "Proc. of RTSS 1995", IEEE Computer Society, 1995, p. 76-89

[51] L. RICKER, K. RUDIE.*Know Means No: Incorporating Knowledge into Discrete-Event Control Systems*, in "IEEE Transactions on Automatic Control", September 2000, vol. 45, n° 9, p. 1656–1668

[52] L. RICKER, K. RUDIE.*Knowledge Is a Terrible Thing to Waste: Using Inference in Discrete-Event Control Problems*, in "IEEE Transactions on Automatic Control", MarchSeptember 2007, vol. 52, n° 3, p. 428–441

[53] C. RODRÍGUEZ, S. SCHWOON, V. KHOMENKO.*Contextual Merged Processes*, in "34th International Conference on Applications and Theory of Petri Nets (ICATPN'13)", Italy, Lecture Notes in Computer Science, Springer, 2013, vol. 7927, p. 29-48 [*DOI :* 10.1007/978-3-642-38697-8_3], https://hal.archives-ouvertes.fr/hal-00926202

[54] H. L. S. YOUNES, R. G. SIMMONS.*Statistical probabilistic model checking with a focus on time-bounded properties*, in "Inf. Comput.", September 2006, vol. 204, p. 1368–1409 [*DOI :* 10.1016/J.IC.2006.05.002], http://dl.acm.org/citation.cfm?id=1182767.1182770

Project-Team OPIS

OPtImization for large Scale biomedical data

IN PARTNERSHIP WITH:
CentraleSupélec

RESEARCH CENTER
Saclay - Île-de-France

THEME
Computational Neuroscience and Medicine

Table of contents

Project-Team OPIS

Creation of the Project-Team: 2019 May 01

Keywords:

Computer Science and Digital Science:

A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.3. - Reinforcement learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A8.2. - Optimization
A8.7. - Graph theory
A9.2. - Machine learning
A9.3. - Signal analysis
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1. - Biology
B1.2. - Neuroscience and cognitive science
B2.6. - Biological and medical imaging

1. Team, Visitors, External Collaborators

Research Scientist

Emilie Chouzenoux [Inria, Researcher, from Sep 2019, HDR]

Faculty Members

Jean-Christophe Pesquet [Team leader, CentraleSupélec, Professor, HDR]
Fragkiskos Malliaros [CentraleSupélec, Associate Professor]
Edouard Oyallon [CentraleSupélec, Associate Professor, until Aug 2019]
Hugues Talbot [CentraleSupélec, Professor, HDR]
Maria Vakalopoulou [CentraleSupélec, Associate Professor]
Marc Castella [Institut Mines-Télécom, Associate Professor]

External Collaborator

Maria Papadomanolaki [CentraleSupélec]

Technical Staff

Nguyen Vu [Inria, Engineer, from Dec 2019]

PhD Students

Samy Ammari [Institut Gustave Roussy, PhD Student]
Enzo Battistella [CentraleSupélec, PhD Student]
Abdulkadir Celikkanat [CentraleSupélec, PhD Student]
Guillaume Chassagnon [Assistance publique/Hôpitaux de Paris, PhD Student, from Jun 2019 until Sep 2019]
Marie-Caroline Corbineau [CentraleSupélec, PhD Student, until Nov 2019]
Tasnim Dardouri [Univ Paris-Nord, PhD Student, from Nov 2019]
Theo Estienne [CentraleSupélec, PhD Student]
Kavya Gupta [Thales, PhD Student, from Nov 2019, granted by CIFRE]
Yunshi Huang [Univ Paris-Est, PhD Student]
Surabhi Jagtap [IFPEN, PhD Student]
Marvin Lerousseau [CentraleSupélec, PhD Student]
Arthur Marmin [CentraleSupélec, PhD Student]
Ana Neacsu [CentraleSupélec, PhD Student, from Oct 2019]
Mihir Sahasrabudhe [CentraleSupélec, PhD Student]
Marion Savanier [GE Healthcare, CentraleSupélec, PhD Student, granted by CIFRE]
Maissa Sghaier [GE Healthcare, CentraleSupélec, PhD Student, granted by CIFRE]
Roger Sun [CentraleSupélec, PhD Student]
Sagar Verma [CentraleSupélec, PhD Student, from Nov 2019]
Mario Viti [GE Healthcare, CentraleSupélec, PhD Student, granted by CIFRE]

Administrative Assistants

Jana Dutrey [CentraleSupélec]
Irina Lahaye [Inria, Administrative Assistant, from Nov 2019]
Alexandra Merlin [Inria, Administrative Assistant, until Sep 2019]

2. Overall Objectives

2.1. OPIS

Mathematical optimization is the key to solving many problems in Science, based on the observation that physical systems obey a general principle of least action. While some problems can be solved analytically, many more can only be solved via numerical algorithms. Research in this domain has been steadily ongoing for decades.

In addition, many fields such as medecine continue to benefit from considerable improvements in data acquisition technology, based on sophisticated tools from optics and physics (e.g., new laser sources in microscopy, multi-coil systems in MRI, novel X-ray schemes in mammography, etc). This evolution is expected to yield significant improvements in terms of data resolution, making the interpretation and analysis of the results easier and more accurate for the practitioner. The large amounts of generated data must be analyzed by sophisticated optimization tools so that, in recent years, optimization has become a main driving force fostering significant advances in data processing. Previously hidden or hard to extract information can be pried from massive datasets by modern recovery and data mining methods. At the same time, automated decision and computer-aided diagnoses are made possible through optimal learning approaches.

However, major bottlenecks still exist. Recent advances in instrumentation techniques come with the need to minimize functions involving increasingly large number of variables (at least one billion variables in 3D digital tomography modality), and with increasingly complex mathematical structure. The computational load for solving these problems may be too high for even state-of-the-art algorithms. New algorithms must be designed with computational scalability, robustness, and versatility in mind. In particular, the following severe requirements must be fulfilled: *i*) ability to tackle high-dimensional problems in a reasonable computation time; *ii*) low-requirements in terms of memory usage; *iii*) robustness to incomplete or unreliable information; *iv*) adaptivity to statistically varying environments; *v*) resilience to latency issues arising in architectures involving multiple computing units.

These difficulties are compounded in the medical and biomedical areas. In these contexts, datasets are not easily available due to patient confidentiality and/or instrument limitations. Moreover, high-level expertise is necessary to interpret the data which can be of very high dimension. Finally, the developed analysis methods must be reliable and interpretable by the medical/biomedical community.

The objective of the OPIS project is to **design advanced optimization methods for the analysis and processing of large and complex data**. Applications to **inverse problems and machine learning tasks in biomedical imaging** will be major outcomes of this research project. We will seek optimization methods able to tackle data with both a large sample-size (“big N" e.g., $N = 10^9$) and/or many measurements (“big P" e.g., $P = 10^4$). The methodologies to be explored will be grounded on nonsmooth functional analysis, fixed point theory, parallel/distributed strategies, and neural networks. The new optimization tools that will be developed will be set in the general framework of graph signal processing, encompassing both regular graphs (e.g., images) and non-regular graphs (e.g., gene regulatory networks).

More specifically, three main research avenues will be explored, namely:

1. proposing novel algorithms able to encompass high-dimensional continuous optimization problems, with established convergence guarantees, and that are well-suited to parallel implementation;
2. designing efficient optimization approaches for the resolution of graph signal processing and graph mining problems;
3. developing a new generation of deep learning strategies, characterized by robustness guarantees, fast training and suitable account for prior information.

Our research program is detailed in the next subsections. We also indicate applications in the medical and biomedical areas on which our program is expected to have a significant impact.

3. Research Program

3.1. Accelerated algorithms for solving high-dimensional continuous optimization problems

Variational problems requiring the estimation of a huge number of variables have now to be tackled, especially in the field of 3D reconstruction/restoration (e.g. $\geq 10^9$ variables in 3D imaging). In addition to the curse of dimensionality, another difficulty to overcome is that the cost function usually reads as the sum of several loss/regularization terms, possibly composed with large-size linear operators. These terms can be nonsmooth and/or nonconvex, as they may serve to promote the sparsity of the sought solution in some suitable representation (e.g. a frame) or to fufill some physical constraints. In such a challenging context, there is a strong need for developing fast parallelized optimization agorithms for which sound theoretical guarantees of convergence can be established. We explore deterministic and stochastic approaches based on proximal tools, MM (Majorization-Minimization) strategies, and trust region methods. Because of the versatility of the methods that will be proposed, a wide range of applications in image recovery are considered: parallel MRI, breast tomosynthesis, 3D ultrasound imaging, and two-photon microscopy. For example, in breast tomosynthesis (collaboration with GE Healthcare), 3D breast images have to be reconstructed from a small number of X-ray projections with limited view angles. Our objective is to facilitate the clinical task by developing advanced reconstruction methods allowing micro-calcifications to be highlighted. In two-photon microscopy (collaboration with XLIM), our objective is to provide effective numerical solutions to improve the 3D resolution of the microscope, especially when cheap laser sources are used, with applications to muscle disease screening.

3.2. Optimization over graphs

Graphs and hypergraphs are rich data structures for capturing complex, possibly irregular, dependencies in multidimensional data. Coupled with Markov models, they constitute the backbones of many techniques used in computer vision. Optimization is omnipresent in graph processing. Firstly, it allows the structure of the underlying graph to be inferred from the observed data, when the former is hidden. Second, it permits to develop graphical models based on the prior definition of a meaningful cost function. This leads to powerful nonlinear estimates of variables corresponding to unknown weights on the vertices and/or the edges of the graph. Tasks such as partitioning the graph into subgraphs corresponding to different clusters (e.g., communities in social networks) or graph matching, can effectively be performed within this framework. Finally, graphs by themselves offer flexible structures for formulating and solving optimization problems in an efficient distributed manner. On all these topics, our group has acquired a long-term expertise that we plan to further strengthen. In terms of applications, novel graph mining methods are proposed for gene regulatory and brain network analysis. For example, we plan to develop sophisticated methods for better understanding the gene regulatory network of various microscopic fungi, in order to improve the efficiency of the production of bio-fuels (collaboration with IFP Energies Nouvelles).

3.3. Toward more understandable deep learning

Nowadays, deep learning techniques efficiently solve supervised tasks in classification or regression by utilizing large amounts of labeled data and the powerful high level features that they learn by using the input data. Their good performance has caught the attention of the optimization community since currently these methods offer virtually no guarantee of convergence, stability or generalization. Deep neural networks are optimized through a computationally intensive engineering process via methods based on stochastic gradient descent. These methods are slow and they may not lead to relevant local minima. Thus, more efforts must be dedicated in order to improve the training of deep neural networks by proposing better optimization algorithms applicable to large-scale datasets. Beyond optimization, incorporating some structure in deep neural networks permits more advanced regularization than the current methods. This should reduce their complexity, as well as allow us to derive some bounds regarding generalization. For example, many signal processing models (e.g. those based on multiscale decompositions) exhibit some strong correspondence with deep learning architectures, yet they do not require as many parameters. One can thus think of introducing some supervision into these models in order to improve their performance on standard benchmarks. A better mathematical understanding of these methods permits to improve them, but also to propose some new models and representations for high-dimensional data. This is particularly interesting in settings such as the diagnosis or prevention of diseases from medical images, because they correspond to critical applications where the made decision is crucial and needs to be interpretable. One of the main applications of this work is to propose robust models for the prediction of the outcome of cancer immunotherapy treatments from multiple and complementary sources of information: images, gene expression data, patient profile, etc (collaboration with Institut Gustave Roussy).

4. Application Domains

4.1. Sparse signal processing in chemistry

Participants: Marc Castella, Emilie Chouzenoux, Arthur Marmin, Jean-Christophe Pesquet (Collaboration: Laurent Duval, IFPEN, Rueil Malmaison)

Mass Spectrometry (MS) is a powerful tool used for robust, accurate, and sensitive detection and quantification of molecules of interest. Thanks to its sensibility and selectivity, MS is widely used in proteomics such anti-doping, metabolomics, medicine or structural biology. In particular, it has applications in clinical research, personalized medicine, diagnosis process and tumours profiling and pharmaceutical quality control. In an MS experiment, the raw signal arising from the molecule ionization in an ion beam is measured as a function of time via Fourier Transform-based measures such as Ion Cyclotron Resonance (FT-ICR) and Orbitrap. A spectral analysis step is then performed to improve the quality of data. The goal is then to determine from this observed pattern distribution the most probable chemical composition of the sample, through the determination of the monoisotopic mass, charge state and abundance of each present molecule. This amounts to solve a large scale signal estimation problem under specific sparsity constraints [35], [55]. Collaboration with Dr. L. Duval, Research Engineer at IFP Energies Nouvelles, France is on-going in this applicative context.

4.2. Image restoration for two-photon microscopy

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Mathieu Chalvidal (Collaboration: Claire Lefort, XLIM, CNRS, Limoges)

Through an ongoing collaboration with physicists from XLIM laboratory (CNRS, Limoges, France), we propose advanced mathematical and computational solutions for multiphoton microscopy (MPM) 3D image restoration. This modality enjoys many benefits such as a decrease in phototoxicity and increase in penetration depth. However, blur and noise issues can be more severe than with standard confocal images. Our objective is to drastically improve the quality of the generated images and their resolution by improving the characterization of the PSF of the system [12] and compensating its effect. We consider the application of the improved MPM imaging tool to the microscopic analysis of muscle ultrastructure and composition, with the aim to help diagnosing muscle disorders including rare and orphan muscle pathologies.

4.3. Representation Learning for Biological Networks

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat (Collaboration: Duong Nguyen, UC San Diego)

Networks (or graphs) are ubiquitous in the domain of biology, as many biological systems can naturally be mapped to graph structures. Characteristic examples include protein-protein interaction and gene regulatory networks. To this extend, machine learning on graphs is an important task with many practical applications in network biology. For example, in the case on protein-protein interaction networks, predicting the function of a protein is a key task that assigns biochemical roles to proteins. The main challenge here is to find appropriate representations of the graph structure, in order to be easily exploited by machine learning models. The traditional approach to the problem was relying on the extraction of "hand-crafted" discriminating features that encode information about the graph, based on user-defined heuristics. Nevertheless, this approach has demonstrated severe limitations, as the learning process heavily depends on the manually extracted features. To this end, feature (or representation) learning techniques can be used to automatically learn to encode the graph structure into low-dimensional feature vectors – which can later be used in learning tasks. Our goal here is to develop a systematic framework for large-scale representation learning on biological graphs. Our approach takes advantage of the clustering structure of these networks, to further enhance the ability of the learned features to capture intrinsic structural properties.

4.4. Breast tomosynthesis

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)

Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing is modeled [50].

4.5. Inference of gene regulatory networks

Participants: Surabhi Jagtap, Fragkiskos Malliaros, Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)

The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in dierent living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

4.6. Imaging biomarkers and characterization for chronic lung diseases

Participants: Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Marie-Pierre Revel and Nikos Paragios: AP-HP - Hopital Cochin Broca Hotel Dieu; Therapanacea)

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i)severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by deep convolutional neural networks and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

4.7. Imaging radiomics and genes to assess immunotherapy

Participants: Samy Ammari, Enzo Batistella, Emilie Chouzenoux, Théo Estienne, Marvin Lerousseau, Hugues Talbot, Roger Sun, Maria Vakalopoulou (in collaboration with Corinne Balleyguier, Caroline Caramella, Éric Deutsch, Nathalie Lassau, Institut de Cancérologie Gustave Roussy, Nikos Paragios, Therapanacea)

Because responses of cancer patients to immunotherapy can vary considerably, innovative predictors of response to treatment are urgently needed to improve patients outcomes.

We have aimed to develop and independently validate a radiomics-based biomarkers of tumour-infiltrating CD8 cells in patients included in phase 1 trials of anti-programmed cell death protein (PD)-1 or anti-programmed cell death ligand 1 (PD-L1) mono-therapy. We also aimed to evaluate the association between the biomarker, tumour immune phenotype and clinical outcomes of these patients.

Concurrently, we have evaluated various ways of estimating patient response to treatment based on well-established radiomics such as estimated tumour count and volumes. Among published metrics, we have select those that shown good predictive power and proposed a new one, which is particularly effective for patient with a poor response [63].

Furthermore, we have developed and validated a novel imaging-based decision-making algorithm for use by the clinician that helps differentiate pituitary metastasis from autoimmune hypophysitis in patients undergoing immune checkpoint blockade therapy [21].

These works are in collaboration with the Institut de Cance´rologie Gustave Roussy Paris.

4.8. Development of a heart ventricle vessel generation model for perfusion analysis

Participant: Hugues Talbot (collaboration with L. Najman ESIEE Paris, I. Vignon-Clementel, REO Team leader, Inria, Charles Taylor, Heartflow Inc.)

Cardio-vascular diseases are the leading cause of mortality in the world. Understanding these diseases is a current, challenging and essential research project. The leading cause of heart malfunction are stenoses causing ischemia in the coronary vessels. Current CT and MRI technology can assess coronary diseases but are typically invasive, requiring catheterization and relatively toxic contrast agents injection. In collaboration with the REO team headed by Irène Vignon-Clementel, and Heartflow, a US based company, we have in the past worked to use image-based exams only, limiting the use of contrast agents and in many cases eliminating catheterisation. Heartflow is current the market leader in non-invasive coronary exams.

Unfortunately, current imaging technology is unable to assess the full length of coronary vessels. CT is limited to a resolution of about 1mm, whereas coronary vessels can be much smaller, down to about 10 micrometers in diameter. Blood perfusion throughout the heart muscle can provide insight regarding coronary health in areas that CT or MRI cannot assess. Perfusion imaging with PET or a Gamma camera, the current gold standard, is an invasive technology requiring the use of radioactive tracers.

We have investigated patient-specific vessel generation models together with porous model simulations in order to propose a forward model of perfusion imaging, based on the known patient data, computer flow dynamic simulations as well as experimental data consistent with known vessel and heart muscle physiology. The objective of this work is to both provide a useful, complex forward model of perfusion image generation, and to solve the inverse problem of locating and assessing coronary diseases given a perfusion exam, even though the affected vessels may be too small to be imaged directly.

In 2019, we have produced a functional myocardial perfusion model consisting of the CT-derived segmented coronary vessels, a simulated vessel tree consisting of several thousands of terminal vessels, filling the myocardium in a patient-specific way, consistent with physiology data, physics-based and empirically-observed vessel growth rules, and a porous medium. We have produced a CFD code capable of simulating blood flow in all three coupled compartments, which allows us to simulate perfusion realistically.

5. Highlights of the Year

5.1. Highlights of the Year

- Our M.Sc. program in Data Sciences and Business Analytics (with ESSEC Business School) was ranked 3rd worldwide in the QS World University Rankings.
- E. Chouzenoux was laureate of the ERC Starting Grant MAJORIS (starting date: 01-01-2020).
- M.C. Corbineau received the best poster award at “Journée de rencontre entre entreprises, doctorants et jeunes docteurs” (J-RED) in 2019.

6. New Software and Platforms

6.1. Platforms

6.1.1. The Proximity Operator Repository

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Giovanni Chierchia, Univ. Paris Est, and Patrick Combettes, North Carolina State University).

link: http://proximity-operator.net/
Proximity operators have become increasingly important tools as basic building blocks of proximal splitting algorithms, a class of algorithms that decompose complex composite convex optimization methods into simple steps involving one of the functions present in the model. This website provides formulas for efficiently computing the proximity operator of various functions, along with the associated codes.

6.1.2. FIGARO

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Claire Lefort, XLIM, Limoges).
link: https://imagejdocu.tudor.lu/plugin/analysis/figaro_psf_3d_optical_microscopy/start
The plugin FIGARO, for ImageJ software, helps tracking resolution of microscope systems by extracting dimensions and orientation of standardized microbeads images, acquired from test samples. In the development of optical microscopes for biomedical imaging, the evaluation of resolution is a fundamental parameter achieved by Point Spread Function (PSF) measurements. Sometimes, PSF measurement procedure is not easy or impossible in case of microspheres images presenting a high noise level. The current method proposed into the plugin FIGARO is based on a variational approach for PSF modeling through multivariate Gaussian fitting, adapted to images acquired in a high noise context [12].

6.1.3. Kymatio

Participant: Edouard Oyallon.
link: http://www.kymat.io
The website shares the software `Kymatio` for Scattering Transform, that computes cascade of wavelets and modulus non-linearity. The codes have been optimized for GPUs and work on the open-source framework *PyTorch*.

6.1.4. BiasedWalk: Learning latent node features with biased random walks

Participants: Fragkiskos Malliaros and Duong Nguyen (UC San Diego)
link: https://github.com/duong18/BiasedWalk/
The BiasedWalk network representation learning algorithm, computes latent node features on graphs based on biased random walks. The framework has been implemented in Python and has been built upon widely used modules, including `networkx`, `numpy`, `gensim` and `scikit-learn`.

6.1.5. DiffuGreedy: Influence maximization in complex networks based on diffusion cascades

Participants: Fragkiskos Malliaros, Georgios Panagopoulos and Michalis Vazirgiannis (École Polytechnique)
link: https://goo.gl/GpfCVZ
The DiffuGreedy is an algorithm for influence maximization in complex networks, that is based on diffusion cascades. It has been implemented in Python and has been built upon widely used modules, including `networkx`, `igraph`, `numpy` and `pandas`.

6.1.6. Graph-based text categorization

Participants: Fragkiskos Malliaros, Konstantinos Skianis and Michalis Vazirgiannis (École Polytechnique)
link: https://github.com/y3nk0/Graph-Based-TC/
Graph-based TC is a framework for text categorization that relies on a graph representation of documents. The framework uses various graph centrality criteria to determine the importance of a term within a document. It also makes use of word embeddings to further boost the performance of graph-based methods. It has been implemented in Python and has been built upon widely used modules, including `networkx`, `igraph`, `numpy` and `scikit-learn`.

6.1.7. KernelNE - Topical Node Embeddings

Participants: Abdulkadir Çelikkanat, Fragkiskos Malliaros
link: https://abdcelikkanat.github.io/projects/kernelNE/

KernelNE learns node representations on graphs based on a weighted matrix factorization model which encodes random walk-based information about the nodes, as described in [53].

6.1.8. EFGE - Exponential Family Graph Embeddings

Participants: Abdulkadir Çelikkanat, Fragkiskos Malliaros
link: https://abdcelikkanat.github.io/projects/EFGE/
EFGE learns node embeddings generalizing random walk-based network representation learning techniques to exponential family conditional distributions, as described in [54].

6.1.9. Semi-supervised Fake News Detection

Participants: Adrien Benamira, Benjamin Devillers, Etienne Lesot, Ayush K. Ray, Manal Saadi, and Fragkiskos Malliaros
link: https://github.com/bdvllrs/misinformation-detection-tensor-embeddings
We propose a graph-based semi-supervised fake news detection method, based on graph neural networks, as described in [34].

6.1.10. The PINK image library

Participant: Hugues Talbot
link: http://ibipio.hu/joomla/
The PINK image library is a general-purpose, open-source, portable image processing library specializing in discrete geometry and mathematical morphology. It is the result of several decades of research in these domains and features state-of-the art algorithmic implementation of both classical and leading edge DG and MM operators. These include nD parallel thinning and skeletonization methods and efficient hierarchical morphological transforms.

6.1.11. The Vivabrain AngioTK toolkit

Participant: Hugues Talbot
link: https://github.com/vivabrain/angiotk
AngioTK is a toolkit supported by Kitware (the authors of VTK) for the filtering, segmentation, generation and simulation of blood vessels. It was started in the context of the Vivabrain ANR project in 2012, but continues with the same as well as new partners. Applications are numerous, from the simulation and understanding of perfusion (see associated theme) to the simulation of realistic blood flow MRI images with associated ground truth, via the generation of blood vessel atlases.

6.1.12. The PET/CT FIJI Viewer

Participant: Hugues Talbot
link: http://petctviewer.org/
Quantitive Positron Emission Tomography is a new area for which software is not readily availble. With our partners from Toulouse and Beth Israel hospitals, we have proposed a free open-source plugin allowing clinicians to view, interact and perform automated and interactive lesion segmentation in the context of PET/CT.

7. New Results

7.1. General risk measures for robust machine learning

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Henri Gérard, ENPC, Paris)

A wide array of machine learning problems are formulated as the minimization of the expectation of a convex loss function on some parameter space. Since the probability distribution of the data of interest is usually unknown, it is is often estimated from training sets, which may lead to poor out-of-sample performance. In the work [11], we bring new insights in this problem by using the framework which has been developed in quantitative finance for risk measures. We show that the original min-max problem can be recast as a convex minimization problem under suitable assumptions. We discuss several important examples of robust formulations, in particular by defining ambiguity sets based on φ-divergences and the Wasserstein metric. We also propose an efficient algorithm for solving the corresponding convex optimization problems involving complex convex constraints. Through simulation examples, we demonstrate that this algorithm scales well on real data sets.

7.2. Deep Latent Factor Model for Collaborative Filtering

Participants: Emilie Chouzenoux (Collaboration: Aanchal Mongia, Neha Jhamb, Angshul Majumdar, IIIT Delhi, India)

Latent factor models have been used widely in collaborative filtering based recommender systems. In recent years, deep learning has been successful in solving a wide variety of machine learning problems. Motivated by the success of deep learning, we propose in [44], [23] a deeper version of latent factor model. Experiments on benchmark datasets shows that our proposed technique significantly outperforms all stateof-the-art collaborative filtering techniques.

7.3. A Proximal Interior Point Algorithm with Applications to Image Processing

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet

In the work [10], we introduce a new proximal interior point algorithm (PIPA). This algorithm is able to handle convex optimization problems involving various constraints where the objective function is the sum of a Lipschitz differentiable term and a possibly nonsmooth one. Each iteration of PIPA involves the minimization of a merit function evaluated for decaying values of a logarithmic barrier parameter. This inner minimization is performed thanks to a finite number of subiterations of a variable metric forward-backward method employing a line search strategy. The convergence of this latter step as well as the convergence the global method itself are analyzed. The numerical efficiency of the proposed approach is demonstrated in two image processing applications.

7.4. Deep Unfolding of a Proximal Interior Point Method for Image Restoration

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet (Collaboration: Carla Bertocchi, Marco Prato, Universita di Modena, Italy)

Variational methods are widely applied to ill-posed inverse problems for they have the ability to embed prior knowledge about the solution. However, the level of performance of these methods significantly depends on a set of parameters, which can be estimated through computationally expensive and time consuming methods. In contrast, deep learning offers very generic and efficient architectures, at the expense of explainability, since it is often used as a black-box, without any fine control over its output. Deep unfolding provides a convenient approach to combine variational-based and deep learning approaches. Starting from a variational formulation for image restoration, we developped in [36], [5], iRestNet, a neural network architecture obtained by unfolding a proximal interior point algorithm. Hard constraints, encoding desirable properties for the restored image, are incorporated into the network thanks to a logarithmic barrier, while the barrier parameter, the stepsize, and the penalization weight are learned by the network. We derive explicit expressions for the gradient of the proximity operator for various choices of constraints, which allows training iRestNet with gradient descent and backpropagation. In addition, we provide theoretical results regarding the stability of the network for a common inverse problem example. Numerical experiments on image deblurring problems show that the proposed approach compares favorably with both state-of-the-art variational and machine learning methods in terms of image quality.

7.5. Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet (Collaboration: Denis Kouamé, Jean-Yves Tourneret, IRIT, Toulouse)
Joint deconvolution and segmentation of ultrasound images is a challenging problem in medical imaging. By adopting a hierarchical Bayesian model, we propose in [15] an accelerated Markov chain Monte Carlo scheme where the tissue reflectivity function is sampled thanks to a recently introduced proximal unadjusted Langevin algorithm. This new approach is combined with a forward-backward step and a preconditioning strategy to accelerate the convergence, and with a method based on the majorization-minimization principle to solve the inner nonconvex minimization problems. As demonstrated in numerical experiments conducted on both simulated and in vivo ultrasound images, the proposed method provides high-quality restoration and segmentation results and is up to six times faster than an existing Hamiltonian Monte Carlo method.

7.6. A Random Block-Coordinate Douglas-Rachford Splitting Method with Low Computational Complexity for Binary Logistic Regression

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Giovanni Chierchia, ESIEE Paris, Luis Bricenos Arias, Universidad Técnica Federico Santa Maria, Valparaiso, Chile)
In the paper [6], we proposed a new optimization algorithm for sparse logistic regression based on a stochastic version of the Douglas-Rachford splitting method. Our algorithm sweeps the training set by randomly selecting a mini-batch of data at each iteration, and it allows us to update the variables in a block coordinate manner. Our approach leverages the proximity operator of the logistic loss, which is expressed with the generalized Lambert W function. Experiments carried out on standard datasets demonstrate the efficiency of our approach w.r.t. stochastic gradient-like methods.

7.7. A probabilistic incremental proximal gradient method

Participant: Emilie Chouzenoux (Collaboration: Omer Deniz Akyildiz, Alan Turing Institute, London, UK, Victor Elvira, University of Edinburgh, Joaquin Miguez, Universidad Carlos III de Madrid, Spain)
In the paper [3], we proposed a probabilistic optimization method, named probabilistic incremental proximal gradient (PIPG) method, by developing a probabilistic interpretation of the incremental proximal gradient algorithm. We explicitly model the update rules of the incremental proximal gradient method and develop a systematic approach to propagate the uncertainty of the solution estimate over iterations. The PIPG algorithm takes the form of Bayesian filtering updates for a state-space model constructed by using the cost function. Our framework makes it possible to utilize well-known exact or approximate Bayesian filters, such as Kalman or extended Kalman filters, to solve large scale regularized optimization problems.

7.8. Optimal Multivariate Gaussian Fitting with Applications to PSF Modeling in Two-Photon Microscopy Imaging

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Claire Lefort, XLIM, Limoges, Tim Tsz-Kit Lau, Northwestern University, USA)

Fitting Gaussian functions to empirical data is a crucial task in a variety of scientific applications, especially in image processing. However, most of the existing approaches for performing such fitting are restricted to two dimensions and they cannot be easily extended to higher dimensions. Moreover, they are usually based on alternating minimization schemes which benefit from few theoretical guarantees in the underlying nonconvex setting. In the paper [12], we provided a novel variational formulation of the multivariate Gaussian fitting problem, which is applicable to any dimension and accounts for possible non-zero background and noise in the input data. The block multiconvexity of our objective function leads us to propose a proximal alternating method to minimize it in order to estimate the Gaussian shape parameters. The resulting FIGARO algorithm is shown to converge to a critical point under mild assumptions. The algorithm shows a good robustness when tested on synthetic datasets. To demonstrate the versatility of FIGARO, we also illustrate its excellent performance in the fitting of the Point Spread Functions of experimental raw data from a two-photon fluorescence microscope.

7.9. Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization

Participants: Emilie Chouzenoux, Loubna El Gueddari (Collaboration: Philippe Ciuciu, Alexandre Vignaut, Inria Saclay, Parietal)

Over the last decade, the combination of parallel imaging (PI) and compressed sensing (CS) in magnetic resonance imaging (MRI) has allowed to speed up acquisition while maintaining a good signal-to-noise ratio (SNR) for millimetric resolution. Self-calibrating techniques such as L1-ESPiRIT have emerged as a standard approach to estimate the coil sensitivity maps that are required at the reconstruction stage. Although straightforward in Cartesian acquisitions, these approaches become more computationally demanding in non-Cartesian scenarios especially for high resolution imaging (e.g. 500 μm in plane). Instead, calibration-less techniques no longer require this prior knowledge to perform multi-channel image reconstruction from undersampled k-space data. In this work, we introduce a new calibration-less PI-CS reconstruction method that is particularly suited to non-Cartesian data. It leverages structure sparsity of the multi-channel images in a wavelet transform domain while adapting to SNR inhomogeneities across receivers thanks to the OSCAR-norm regularization. Comparison and validation on 8 to 20-fold prospectively accelerated high-resolution ex-vivo human brain MRI data collected at 7 Tesla shows that the subbandwise OSCAR-norm regularization achieves the best trade-off between image quality and computational cost at the reconstructions stage compared to other tested versions (global, scalewise and pixelwise). This approach provides slight to moderate improvement over its state-of-the-art competitors (self-calibrating ‘1-ESPIRiT method and calibration-less AC-LORAKS and CaLM methods) in terms of closeness to the Cartesian reference magnitude image. Importantly, it also preserves much better phase information compared to other approaches [37], [57], [62].

7.10. Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Alessandro Benfenati, Universita di Milano)

In recent years, there has been a growing interest in mathematical models leading to the minimization, in a symmetric matrix space, of a Bregman divergence coupled with a regularization term. We address problems of this type within a general framework where the regularization term is split into two parts, one being a spectral function while the other is arbitrary. A Douglas–Rachford approach is proposed to address such problems, and a list of proximity operators is provided allowing us to consider various choices for the fit–to–data functional and for the regularization term. Based on our theoretical results, two novel approaches are proposed for the noisy graphical lasso problem, where a covariance or precision matrix has to be statistically estimated in presence of noise. The Douglas–Rachford approach directly applies to the estimation of the covariance matrix. When the precision matrix is sought, we solve a nonconvex optimization problem. More precisely, we propose a majorization–minimization approach building a sequence of convex surrogates and solving the inner optimization subproblems via the aforementioned Douglas–Rachford procedure. We establish conditions for the convergence of this iterative scheme. We illustrate the good numerical performance of the proposed approaches with respect to state–of–the–art approaches on synthetic and real-world datasets [4].

7.11. Representation Learning on Real-World Graphs

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat

Network representation learning (NRL) methods aim to map each vertex into a low dimensional space by preserving both local and global structure of a given network. In recent years, various approaches based on random walks have been proposed to learn node embeddings – thanks to their success in several challenging problems. In this work, we have introduced two methodologies to compute latent representations of nodes based on random walks.

In particular, we have proposed *Kernel Node Embeddings* (KernelNE) [53], a model that aims to bring together two popular approaches for NRL, namely matrix factorization and random walk-based models. KernelNE is a weighted matrix factorization model which encodes random walk-based information about the nodes of the graph. The main benefit of this formulation is that it allows to utilize kernel functions on the computation of the embeddings.

Our second approach is motivated by the fact that the popular Skip-Gram algorithm models the conditional distribution of nodes within a random walk based on the softmax function, which might prohibit to capture richer types of interaction patterns among nodes that co-occur within a random walk. Here we argue that considering more expressive conditional probability models to relate nodes within a random walk sequence, might lead to more informative representations. That way, we have introduced the *Exponential Family Graph Embedding* (EFGE) model [54], that capitalizes on exponential family distribution models to capture interactions between nodes.

We have evaluated our methods on two downstream tasks: node classification and link prediction in social, information and biological networks. The experimental results demonstrate that random walk-based models accompanied with kernels as well as exponential family distributions outperform widely-known baseline NRL methods.

7.12. Semi-supervised Learning for Misinformation Detection

Participants: Fragkiskos Malliaros (Collaboration: Adrien Benamira, Benjamin Devillers, Etienne Lesot, Ayush K. Ray, Manal Saadi, CentraleSupélec)

Social networks have become the main platforms for information dissemination. Nevertheless, due to the increasing number of users, social media platforms tend to be highly vulnerable to the propagation of disinformation – making the detection of fake news a challenging task. In our work, we have focused on content-based methods for detecting fake news – casting the problem to a binary text classification one (an article corresponds to either fake news or not). The main challenge here stems from the fact that the number of labeled data is limited; very few articles can be examined and annotated as fake. To this extend, we opted for semi-supervised learning approaches. In particular, we have proposed a graph-based semi-supervised fake news detection method, based on graph neural networks [34]. Our intuition is that, graphs are expressive models that are able to capture contextual dependencies among articles, alleviating the label scarcity constraint. On a high level, our framework is composed of three components: i) embedding of articles in the Euclidean space; ii) construction of an article similarity graph; iii) inference of missing labels using graph learning techniques. The experimental results indicate that the proposed methodology achieves better performance compared to traditional classification techniques, especially when trained on limited number of labeled articles.

7.13. A Perturb and Combine Approach to Analyze Real-World Graphs

Participants: Fragkiskos Malliaros (Collaboration: Antoine J.-P. Tixier, Maria Evgenia G. Rossi, Jesse Read, Michalis Vazirgiannis, École Polytechnique)

Influential spreaders are nodes that can diffuse information to the largest part of the network in a minimum amount of time. Detecting influential spreaders is an important task with numerous real-world applications. Nevertheless, some of the most effective influential spreader detection algorithms (e.g., the k-core decomposition) are unstable to small perturbations of the network structure. Inspired by bagging in Machine Learning, we have proposed the first Perturb and Combine (P&C) procedure for networks [51]. It (1) creates many perturbed versions of a given graph, (2) applies a node scoring function separately to each graph, and (3) combines the results. Experiments conducted on real-world networks of various sizes with the k-core, generalized k-core, and PageRank algorithms reveal that P&C brings substantial improvements. Moreover, this performance boost can be obtained at almost no extra cost through parallelization. Finally, a bias-variance analysis suggests that P&C works mainly by reducing bias, and that therefore, it should be capable of improving the performance of all vertex scoring functions, including stable ones.

7.14. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping: Mean-square and linear convergence

Participant: Jean-Christophe Pesquet (Collaboration: Patrick Louis Combettes, North Carolina University, USA)

In our previous work, we investigated the almost sure weak convergence of block-coordinate fixed point algorithms and discussed their applications to nonlinear analysis and optimization. This algorithmic framework features random sweeping rules to select arbitrarily the blocks of variables that are activated over the course of the iterations and it allows for stochastic errors in the evaluation of the operators. The paper [14] establishes results on the mean-square and linear convergence of the iterates. Applications to monotone operator splitting and proximal optimization algorithms are presented.

7.15. Rational optimization for non- linear reconstruction with approximate ℓ_0 penalization

Participants: Marc Castella, Arthur Marmin, Jean-Christophe Pesquet

Recovering nonlinearly degraded signal in the presence of noise is a challenging problem. In this work, this problem is tackled by minimizing the sum of a non convex least-squares fit criterion and a penalty term. We assume that the nonlinearity of the model can be accounted for by a rational function. In addition, we suppose that the signal to be sought is sparse and a rational approximation of the ℓ_0 pseudo-norm thus constitutes a suitable penalization. The resulting composite cost function belongs to the broad class of semi-algebraic functions. To find a globally optimal solution to such an optimization problem, it can be transformed into a generalized moment problem, for which a hierarchy of semidefinite programming relaxations can be built. Global optimality comes at the expense of an increased dimension and, to overcome computational limitations concerning the number of involved variables, the structure of the problem has to be carefully addressed. A situation of practical interest is when the nonlinear model consists of a convolutive transform followed by a componentwise nonlinear rational saturation. We then propose to use a sparse relaxation able to deal with up to several hundreds of optimized variables. In contrast with the naive approach consisting of linearizing the model, our experiments show that the proposed approach offers good performance [7].

7.16. Deep neural network structures solving variational inequalities

Participant: Jean-Christophe Pesquet (Collaboration: Patrick Louis Combettes, North Carolina University, USA)

Motivated by structures that appear in deep neural networks, we investigate nonlinear composite models alternating proximity and affine operators defined on different spaces. We first show that a wide range of activation operators used in neural networks are actually proximity operators. We then establish conditions for the averagedness of the proposed composite constructs and investigate their asymptotic properties. It is shown that the limit of the resulting process solves a variational inequality which, in general, does not derive from a minimization problem [13].

7.17. Generation of patient-specific cardiac vascular networks: a hybrid image-based and synthetic geometric model

Participant: Hugues Talbot (Collaboration: Clara Jaquet, Laurent Najman, ESIEE Paris, Leo Grady, Michiel Schaap, Buzzy Spain, Hyun Kim, Charles Taylor, HeartFlow Inc, Irene Vignon-Clementel, Inria Paris)
In this work, we have proposed a blood-vessel generation procedure for extending known patient vasculature over and within the heart ventricle [19]. It is patient-specific, in the sense that it extend the known, segmented patient vasculature, and it is consistent with physics-based blood vessels characteristics (i.e. derived from CFD) and known vessel physiology. The generated vascular network bridges the gap between the vasculature that can be imaged and assessed via classical means (CT or MRI) and perfusion maps that can be imaged with specific modalities (radiotracer injected scintigraphy or PET). One objective of this work is to eventually propose a forward model for perfusion map generation, that can be used to solved the associated inverse problem of finding the cause of observed perfusion deficits associated with coronary diseases that cannot be imaged directly.

7.18. High throughput automated detection of axial malformations in Medaka fish embryo

Participant: Hugues Talbot (Collaboration: Diane Genest, Élodie Puybareau, Jean Cousty, ESIEE Paris, Marc Léonard, Noémie de Crozé, L'Oréal Recherche)
Fish embryos are used throughout the cosmetics industry to assess the toxicity of the components of their products, as well as more generally in waterways pollution measurements. Indeed pollution is often detectable in trace amounts when they hinder, stop or cause malformations during fish embryo development. In this work, we propose a high-throughput procedure for detecting most important malformations in fish embryo. For examples those affecting the tail or the eyes, based on image analysis and machine learning [16]. We have also proposed an atlas-based automated procedure for detecting swim bladder malformations, which are very difficult to assess manually [39].

These malformation are among the most difficult to assess but very common in various degrees of severity. Our procedure provide similar error rate as trained and careful humans operators, as assessed on thousands of images acquired in partneship with L'Oréal. We also show that our procedure is much faster and more consistent than human operators. It is now used in production by our partner.

7.19. Quantitative PET in the context of lymphoma

Participant: Hugues Talbot (Collaboration: Eloïse Grossiord, Laurent Najman, ESIEE Paris, Benoît Naegel, iCube, Strasbourg, Nicolas Passat, CRESTIC, Reims)
Lymphoma is type of cancer affecting the lymph system. Similar to blood disorders, these cancers can be difficult to cure because they affect a large portion of the body and metastasize easily. In contrast to leukemia, lymphoma also affects organs: the lymph nodes. Assessing the effectiveness of therapies implies to follow the impact of treatment on lymph nodes. This requires segmenting a large number of lesion, often several dozens. In [17], we have proposed an automated procedure based on hierarchical mathematical morphology, which has been extensively validated, and is now available as a plug-in for ImageJ/FIJI.

7.20. nD variational restoration of curvilinear structures with prior-based directional regularization

Participant: Hugues Talbot (Collaboration: Odyssée Merveille, Benoît Naegel, iCube, Strasbourg, Nicolas Passat, CRESTIC, Reims)

Curvilinear structure restoration in image processing procedures is a difficult task, which can be compounded when these structures are thin, i.e., when their smallest dimension is close to the resolution of the sensor. Many recent restoration methods involve considering a local gradient-based regularization term as prior, assuming gradient sparsity. An isotropic gradient operator is typically not suitable for thin curvilinear structures, since gradients are not sparse for these. In this paper [22], we propose a mixed gradient operator that combines a standard gradient in the isotropic image regions, and a directional gradient in the regions where specific orientations are likely. In particular, such information can be provided by curvilinear structure detectors (e.g., RORPO or Frangi filters). Our proposed mixed gradient operator, that can be viewed as a companion tool of such detectors, is proposed in a discrete framework and its formulation/computation holds in any dimension; in other words, it is valid in Z^n, $n \geq 1$. We show how this mixed gradient can be used to construct image priors that take edge orientation, as well as intensity, into account, and then involved in various image processing tasks while preserving curvilinear structures. The experiments carried out on 2D, 3D, real, and synthetic images illustrate the relevance of the proposed gradient, and its use in variational frameworks for both denoising and segmentation tasks.

7.21. Skin aging automated assessment

Participant: Hugues Talbot (Collaboration: Julie Robic, Alex Nkengne, Clarins laboratories, Benjamin Perret, Michel Couprie, ESIEE Paris)

With aging, human skin becomes drier, thinner and more irregular, but these characteristics are highly person-dependent, and can be brought about via exposure to heat, cold or Sun. It is important to the cosmetics industry to assess objectively the effect of their products on skin aging. With our partner Clarins laboratory, we have proposed a series of automated procedures based on graph-based image analysis. We have in particular proposed to detect the surface that correspond to the dermal-epidermal junction [25], and a series or procedures to link the appearance of this surfact to aging characteristics [48]. Both have been validated by dermatologists specialized in skin aging.

7.22. Particle tracking

Participant: Hugues Talbot (Collaboration: Alessandro Benfenati, Universita di Milano, Francesco Bonacci, Laboratoire Navier, Tarik Bourouina, ESIEE Paris)

Fluorescent bead tracking is important in biomedical application related to biomechanics, rheology and fluid dynamics. We have made several contributions for the detection and tracking of micrometer-scale fluorescent bead in 3D confocal microscopy [47], [61]. Many software packages exist for 2D tracking but almost none exist for 3D. It is a harder problem because in general beads are not fixed and move between plane acquisitions, due to the relatively slow scanning characteristics of confocal microscopy.

7.23. Artificial Intelligence Applications for Thoracic imaging

Participants: Guillaume Chassagnon, Maria Vakalopoulou (Collaboration: Marie-Pierre Revel and Nikos Paragios, AP-HP - Hopital Cochin Broca Hotel Dieu, Therapanacea)

Relevance and penetration of machine learning in clinical practice is a recent phenomenon with multiple applications being currently under development. Deep learning –and especially convolutional neural networks (CNNs)– is a subset of machine learning, which has recently entered the field of thoracic imaging. The structure of neural networks, organized in multiple layers, allows them to address complex tasks. For several clinical situations, CNNs have demonstrated superior performance as compared with classical machine learning algorithms and in some cases achieved comparable or better performance than clinical experts. Chest radiography, a high-volume procedure, is a natural application domain because of the large amount of stored images and reports facilitating the training of deep learning algorithms. Several algorithms for automated reporting have been developed. The training of deep learning algorithm CT images is more complex due to the dimension, variability, and complexity of the 3D signal. The role of these methods is likely to increase in clinical practice as a complement of the radiologist's expertise. The objective of these two reviews [9], [26] is to provide definitions for understanding the methods and their potential applications for thoracic imaging.

7.24. Use of Elastic Registration in Pulmonary MRI for the Assessment of Pulmonary Fibrosis in Patients with Systemic Sclerosis

Participants: Guillaume Chassagnon, Maria Vakalopoulou (Collaboration: Charlotte Martin, Rafael Marini Silva, Alexis Régent, Luc Mouthon, Nikos Paragios and Marie-Pierre Revel, AP-HP - Hopital Cochin Broca Hotel Dieu, Therapanacea)

Elastic registration of inspiratory and expiratory MRI revealed qualitative and quantitative differences in lung deformation in study participants with systemic sclerosis compared with healthy volunteers. Current imaging methods are not sensitive to changes in pulmonary function resulting from fibrosis. MRI with ultrashort echo time can be used to image the lung parenchyma and lung motion. To evaluate elastic registration of inspiratory to expiratory lung MRI for the assessment of pulmonary fibrosis in study participants with systemic sclerosis (SSc). This prospective study [8] was performed from September 2017 to March 2018 and recruited healthy volunteers and participants with SSc and high-resolution CT (within the previous 3 months) of the chest for lung MRI. Two breath-hold, coronal, three-dimensional, ultrashort–echo-time, gradient-echo sequences of the lungs were acquired after full inspiration and expiration with a 3.0-T unit. Images were registered from inspiration to expiration by using an elastic registration algorithm. Jacobian determinants were calculated from deformation fields and represented on color maps. Similarity between areas with marked shrinkage and logarithm of Jacobian determinants were compared between healthy volunteers and study participants with SSc. Receiver operating characteristic curve analysis was performed to determine the best Dice similarity coefficient threshold for diagnosis of fibrosis. Sixteen participants with SSc (seven with pulmonary fibrosis at high-resolution CT) and 11 healthy volunteers were evaluated. Areas of marked shrinkage during expiration with logarithm of Jacobian determinants less than -0.15 were found in the posterior lung bases of healthy volunteers and in participants with SSc without fibrosis, but not in participants with fibrosis. The sensitivity and specificity of MRI for presence of fibrosis at high-resolution CT were 86% and 75%, respectively (area under the curve, 0.81; P = .04) by using a threshold of 0.36 for Dice similarity coefficient. Elastic registration of inspiratory to expiratory MRI shows less lung base respiratory deformation in study participants with systemic sclerosis related pulmonary fibrosis compared with participants without fibrosis.

7.25. U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets

Participants: Théo Estienne, Enzo Battistella, Marvin Lerousseau, Roger Sun, Maria Vakalopoulou (Collaboration: Stergios Christodoulidis, Alexandre Carre, Guillaume Klausner, Stavroula Mougiakakou, Charlotte Robert, Nikos Paragios and Eric Deutsch, Institute Gustave Roussy, University of Bern, Therapanacea)

We proposed in [58] a 3D deep neural network called U-ReSNet, a joint framework that can accurately register and segment medical volumes. The proposed network learns to automatically generate linear and elastic deformation models, trained by minimizing the mean square error and the local cross correlation similarity metrics. In parallel, a coupled architecture is integrated, seeking to provide segmentation maps for anatomies or tissue patterns using an additional decoder part trained with the dice coefficient metric. U-ReSNet is trained in an end to end fashion, while due to this joint optimization the generated network features are more informative leading to promising results compared to other deep learning-based methods existing in the literature. We evaluated the proposed architecture using the publicly available OASIS 3 dataset, measuring the dice coefficient metric for both registration and segmentation tasks. Our promising results indicate the potentials of our method which is composed from a convolutional architecture that is extremely simple and light in terms of parameters.

7.26. Gene Expression High-Dimensional Clustering Towards a Novel, Robust, Clinically Relevant and Highly Compact Cancer Signature

Participants: Enzo Battistella, Théo Estienne, Marvin Lerousseau, Roger Sun, Maria Vakalopoulou (Collaboration: Charlotte Robert, Nikos Paragios and Eric Deutsch, Institute Gustave Roussy, Therapanacea)

Precision medicine, a highly disruptive paradigm shift in healthcare targeting the personalizing treatment, heavily relies on genomic data. However, the complexity of the biological interactions, the important number of genes as well as the lack of substantial patient's clinical data consist a tremendous bottleneck on the clinical implementation of precision medicine. In this work [32], we introduce a generic, low dimensional gene signature that represents adequately the tumor type. Our gene signature is produced using LP-stability algorithm, a high dimensional center-based unsupervised clustering algorithm working in the dual domain, and is very versatile as it can consider any arbitrary distance metric between genes. The gene signature produced by LP-stability reports at least 10 times better statistical significance and 35% better biological significance than the ones produced by two referential unsupervised clustering methods. Moreover, our experiments demonstrate that our low dimensional biomarker (27 genes) surpass significantly existing state of the art methods both in terms of qualitative and quantitative assessment while providing better associations to tumor types than methods widely used in the literature that rely on several omics data.

7.27. A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks

Participants: Maria Papadomanolaki and Maria Vakalopoulou (Collaboration: Konstantinos Karantzalos, National Technical University of Athens)

Deep learning architectures have received much attention in recent years demonstrating state-of-the-art performance in several segmentation, classification and other computer vision tasks. Most of these deep networks are based on either convolutional or fully convolutional architectures. In this study [24], we propose a novel object-based deep-learning framework for semantic segmentation in very high-resolution satellite data. In particular, we exploit object-based priors integrated into a fully convolutional neural network by incorporating an anisotropic diffusion data preprocessing step and an additional loss term during the training process. Under this constrained framework, the goal is to enforce pixels that belong to the same object to be classified at the same semantic category. We compared thoroughly the novel object-based framework with the currently dominating convolutional and fully convolutional deep networks. In particular, numerous experiments were conducted on the publicly available ISPRS WGII/4 benchmark datasets, namely Vaihingen and Potsdam, for validation and inter-comparison based on a variety of metrics. Quantitatively, experimental results indicate that, overall, the proposed object-based framework slightly outperformed the current state-of-the-art fully convolutional networks by more than 1% in terms of overall accuracy, while intersection over union results are improved for all semantic categories. Qualitatively, man-made classes with more strict geometry such as buildings were the ones that benefit most from our method, especially along object boundaries, highlighting the great potential of the developed approach.

7.28. A multi-task deep learning framework coupling semantic segmentation and image reconstruction for very high resolution imagery

Participants: Maria Papadomanolaki and Maria Vakalopoulou (Collaboration: Konstantinos Karantzalos, National Technical University of Athens)

Semantic segmentation, especially for very high-resolution satellite data, is one of the pillar problems in the remote sensing community. Lately, deep learning techniques are the ones that set the state-of-the-art for a number of benchmark datasets, however, there are still a lot of challenges that need to be addressed, especially in the case of limited annotations. To this end, in this study [45], we propose a novel framework based on deep neural networks that is able to address concurrently semantic segmentation and image reconstruction in an end to end training. Under the proposed formulation, the image reconstruction acts as a regularization, constraining efficiently the solution in the entire image domain. This self-supervised component helps significantly the generalization of the network for the semantic segmentation, especially in cases of a low number of annotations. Experimental results and the performed quantitative evaluation on the publicly available ISPRS (WGIII/4) dataset indicate the great potential of the developed approach.

7.29. Detecting Urban Changes with Recurrent Neural Networks from Multitemporal Sentinel-2 Data

Participants: Maria Papadomanolaki, Sagar Verma and Maria Vakalopoulou (Collaboration: Siddharth Gupta and Konstantinos Karantzalos, Granular AI and National Technical University of Athens)

The advent of multitemporal high resolution data, like the Copernicus Sentinel-2, has enhanced significantly the potential of monitoring the earth's surface and environmental dynamics. In this study [45], we present a novel deep learning framework for urban change detection which combines state-of-the-art fully convolutional networks (similar to U-Net) for feature representation and powerful recurrent networks (such as LSTMs) for temporal modeling. We report our results on the recently publicly available bi-temporal Onera Satellite Change Detection (OSCD) Sentinel-2 dataset, enhancing the temporal information with additional images of the same region on different dates. Moreover, we evaluate the performance of the recurrent networks as well as the use of the additional dates on the unseen test-set using an ensemble cross-validation strategy. All the developed models during the validation phase have scored an overall accuracy of more than 95%, while the use of LSTMs and further temporal information, boost the F1 rate of the change class by an additional 1.5%.

7.30. Image Registration of Satellite Imagery with Deep Convolutional Neural Networks

Participants: Maria Vakalopoulou and Mihir Sahasrabudhe (Collaboration: Stergios Christodoulidis, Stavroula Mougiakakou and Nikos Paragios, University of Bern and Therapanacea)

Image registration in multimodal, multitemporal satellite imagery is one of the most important problems in remote sensing and essential for a number of other tasks such as change detection and image fusion. In this study [52], inspired by the recent success of deep learning approaches we propose a novel convolutional neural network architecture that couples linear and deformable approaches for accurate alignment of remote sensing imagery. The proposed method is completely unsupervised, ensures smooth displacement fields and provides real time registration on a pair of images. We evaluate the performance of our method using a challenging multitemporal dataset of very high resolution satellite images and compare its performance with a state of the art elastic registration method based on graphical models. Both quantitative and qualitative results prove the high potentials of our method.

7.31. Lifting AutoEncoders: Unsupervised Learning of a Fully-Disentangled 3D Morphable Model Using Deep Non-Rigid Structure From Motion

Participants: Mihir Sahasrabudhe (Collaboration: Eqward Bartrum, Riza Alp Guler, Dimitris Samaras and Iasonas Kokkinos, Stony Brook, UCL and Ariel AI)

We introduced, in [49], Lifting Autoencoders, a generative 3D surface-based model of object categories. We bring together ideas from non-rigid structure from motion, image formation, and morphable models to learn a controllable, geometric model of 3D categories in an entirely unsupervised manner from an unstructured set of images. We exploit the 3D geometric nature of our model and use normal information to disentangle appearance into illumination, shading, and albedo. We further use weak supervision to disentangle the non-rigid shape variability of human faces into identity and expression. We combine the 3D representation with a differentiable renderer to generate RGB images and append an adversarially trained refinement network to obtain sharp, photorealistic image reconstruction results. The learned generative model can be controlled in terms of interpretable geometry and appearance factors, allowing us to perform photorealistic image manipulation of identity, expression, 3D pose, and illumination properties.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- PhD Contract with IFP Energies Nouvelles

  Project title: Polynomial optimization for sparse signal recovery

  Duration: 2018-2020

  Leaders: M. Castella and J.-C. Pesquet
- PhD Contract with IFP Energies Nouvelles

  Project title: Seismic signal analysis by using neural networks

  Duration: 2019-2022

  Leaders: A. Fraysse and J.-C. Pesquet
- PhD Contract with Thales Group

  Project title: Neural network solutions for safety of complex systems

  Duration: 2019-2022

  Responsible: J.-C. Pesquet and F. Malliaros
- PhD Contract with General Electric Healthcare

  Project title: Minimally invasive assessement of coronary disease

  Duration: 2018-2021

  Leader: Hugues Talbot
- PhD Contract with General Electric Healthcare

  Project title: Optimization methods for breast tomosynthesis

  Duration: 2017-2020

  Leader: J.-C. Pesquet and E. Chouzenoux
- PhD Contract with General Electric Healthcare

  Project title: Reconstruction 3D interventionnelle

  Duration: 2019-2022

  Leader: J.-C. Pesquet and E. Chouzenoux
- PhD Contract with IFP Energies nouvelles

  Project title: Graph-based learning from integrated multi-omics and multi-species data

  Duration: 2019-2022

  Leader: F. Malliaros and J.-C. Pesquet
- Contract with Schneider Electric

  Project title: Neural network modeling of electrical motors

  Duration: 2019

  Leader: J.-C. Pesquet
- Contract with SNCF

  Project title: SIARA project: Developing an automatic system based on deep learning which monitors different types of defects in the current railway network of France.

  Duration: 2018-2019

  Leader: M. Vakalopoulou
- Contract with SNCF

Project title: SNCF Platipus: Examining the potential of machine learning algorithms in the analysis of scouring reports of aquatic foundations.

Duration: 2019-2020

Leader: F. Malliaros, M. Vakalopoulou.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- DATAIA UltraBioLearn (2019-2022). The project aims to research machine learning approaches for medical applications, in particular by leveraging semi-supervised learning using generative, graph-based and certifiable networks, in the context of predicting patient response to cancer treatments. Responsible: H. Talbot, F. Malliaros (N. Lassau, Institut Gustave Roussy).

9.2. National Initiatives

9.2.1. ANR

- Program: ANR PRC

  Project acronym: CoMeDIC

  Project title: Convergent Metrics for DIscrete Calculus

  Duration: 2016-2021

  Coordinator: J.-O. Lachaud (Univ. Rhones Alpes Savoie Mont-Blanc), Local: H. Talbot
- Program: ANR PRCE

  Project acronym: R-Vessel-X

  Project title: Extraction et interprétation robustes des réseaux vasculaires dans les images biomédicales hépatiques

  Duration: 2018-2022

  Coordinator: A. Vacavant (Univ. Clermont Auvergne), local: H. Talbot
- Program: ANR JCJC

  Project acronym: MajIC

  Project title: Majorization-Minimization Algorithms for Image Computing

  Duration: 2017-2021

  Coordinator: E. Chouzenoux
- Program: ANR JCJC

  Project acronym: AVENUE

  Project title: A Visual memory network for scene understanding

  Duration: 2018-2022

  Coordinator: Dr. Karteek Alahari (Inria Grenoble - Rhône-Alpes). Local: F. Malliaros.

9.2.2. Others

- Program: CNRS-CEFIPRA

  Project acronym: NextGenBP

  Project title: Looking Beyond Backpropagation in Deep Learning

  Duration : 2017-2019

  Coordinator: E. Chouzenoux

- Program: PHC - Campus France
  Projet acronym: POLONIUM
  Project title: When Poisson and Gauss meet in imaging
  Duration: 2018-2020
  Coordinator: J.C. Pesquet

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

- Program: H2020 ITN Marie Sklodowska-Curie
  Projet acronym: SUNDIAL
  Project Title: SUrvey Network for Deep Imaging Analysis and Learning
  Duration: 2017-2021
  Coordinator: Reynier Peletier (U. Groningen, NL), local: Hugues Talbot

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

- Sup'Com Tunis - Prof. Amel Benazza-Benhayia. Collaboration Topic: Multispectral imaging and image compression.
- North Carolina State University - Prof. Patrick Louis Combettes. Collaboration Topic: Fixed point theory.
- Heriot-Watt University, UK - Prof. Audrey Repetti and Prof. Yves Wiaux. Collaboration Topic: Large-scale image restoration.
- University of Edinburgh, UK - Prof. Victor Elvira. Collaboration Topic: Bayesian signal processing.
- Indraprastha Information Institute Technology, Delhi, India - Prof. Angshul Majumdar. Collaboration Topic: Dictionary learning.
- Universidad Técnica Federico Santa María, Valparaíso, Chile - Prof. Luis M. Briceño-Arias. Collaboration Topic: Stochastic optimization.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Prof. Angshul Majumdar, IIIT Delhi, India, June 2019 and December 2019
- Prof. Apostolos N. Papadopoulos, Aristotle University of Thessaloniki, June 2019 to July 2019
- Prof. Patrick L. Combettes, North Carolina State University, US, 18-22 February 2019
- W. Tang (PhD student), North Carolina State University, US, 2-29 May 2019
- S. Sharma (PhD student), IIIT Delhi, India, June 2019 to August 2019

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- M. Vakalopoulou, visiting researcher for 1 month (June-July 2019): Stony Brook University, research team of D. Samaras.
- T. Estienne, 4 months internship (May-August 2019): Center for Biomedical Image Computing and Analytics (CBICA) of University of Pennsylvania.

- E. Battistella, 4 months internship (August-December 2019): Computational Robotics, AI & Biomedicine Lab of RICE University.
- M. Sahasrabudhe, 2 months intership (November-December 2019): Boston Children's Hospital & Harvard Medical School.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- F. Malliaros. Organizer of the Summer School in Artificial Intelligence, CentraleSupélec, France, 2019.

10.1.1.2. Member of the Organizing Committees

- H. Talbot. Member of the organizing committee for ICCV 2023 in Paris.
- F. Malliaros. Member of the organizing committee of the International Workshop of Deep Learning for Graphs and Structured Data Embedding (DL4G-SDE), The Web Conference (WWW), San Francisco, California, 2019.
- F. Malliaros. Member of the organizing committee for the special session on Machine Learning with Graphs: Algorithms and Applications, 28th International Conference on Artificial Neural Networks (ICANN), Munich, Germany, 2019.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- E. Chouzenoux. Elected member of the EURASIP Signal and Data Analytics for Machine Learning technical committee (since 2018), and of the IEEE Signal Processing Theory and Methods technical committee (since 2017).

10.1.2.2. Member of the Conference Program Committees

- E. Chouzenoux. Organization of a special session at the 45th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2020.
- M. Castella. Organization of a special session at the 27th European Signal Processing Conference, EUSIPCO 2019.
- F. Malliaros. Member of the program committee at International Conference on Artificial Neural Networks, ICANN 2019.
- J.C. Pesquet. Member of the program committee at GRETSI 2019.
- H. Talbot. Member of the board and program committee at the International Symposium in Mathematical Morphology, which took place in Saarbrücken, Germany, July 2019.
- H. Talbot. Member of the program committee at the conference Discrete Geometry for Computer Imagery, which took place in Noisy-le-Grand, France, March 2019.
- M. Vakalopoulou. Member of the program committee at the Joint Urban Remote Sensing Event (JURSE), 2019.

10.1.2.3. Reviewer

The members of the team reviewed numerous papers for several international conferences, such as for the annual conferences on Computer Vision and Pattern Recognition (CVPR), Medical Image Computing and Computer Assisted Intervention (MICCAI), Neural Information Processing Systems (NIPS) and International Conference on Learning Representations (ICLR), IEEE International Conference and Acoustics Speech and Signal Processing (ICASSP), IEEE International Conference on Image Processing (ICIP), IEEE Statistical Signal Processing workshop (SSP), European Signal Processing Conference (EUSIPCO), AAAI Conference on Artificial Intelligence (AAAI), The Web Conference (WWW), Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), International Conference on Web and Social Media (ICWSM), International Conference on Machine Learning (ICML), Conference on Neural Information Processing Systems (NeurIPS), International Conference on Complex Networks and Their Applications (Complex Networks), International Workshop on Graph-Based Natural Language Processing (TextGraphs).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Emilie Chouzenoux: Associate Editor at IEEE Transactions on Signal Processing
- Jean-Christophe Pesquet: Associate Editor of the SIAM Journal on Imaging Sciences
- Hugues Talbot: Editor in Chief, Mathematical Morphology, Theory and Applications (De Gruyter).
- Hugues Talbot: Senior area editor for IEEE Signal Processing Letters.

10.1.3.2. Reviewer - Reviewing Activities

- E. Chouzenoux: IEEE Trans. on Image Processing, IEEE Trans. Signal Processing, SIAM Journal on Imaging Science, Journal of Optimization Theory and Applications, Journal of Global Optimization, Journal of Mathematical Imaging and Vision, Inverse Problems.
- J.-C. Pesquet: IEEE Trans. on Signal Processing, IEEE Trans. on Image Processing, IEEE Trans. on Information Theory (IEEE-TI), Signal Processing, SIAM Journal on Optimization, SIAM Journal on Imaging Sciences, Journal of Mathematical Imaging and Vision, Journal of Optimization Theory and Applications, Mathematical Programming.
- F. Malliaros: IEEE Transactions on Knowledge and Data Engineering (TKDE), Applied Network Science, International Journal on Artificial Intelligence Tools (IJAIT).
- H. Talbot: IEEE Pattern Analysis and Machine Intelligence, IEEE Transaction on Image Processing, Computer Vision and Image Understanding, J. on Mathematical Imaging and Vision, Signal Processing Letters, Transactions on Signal Processing. Reviewer for the conferences ACCV, BMVC, CVPR, ICIP, ICML, ISBI, and NeurIPS (ex-NIPS),.
- M. Vakalopoulou: International Journal of Computer Assisted Radiology and Surgery (IJCARS), IEEE Trans. on Geoscience and Remote Sensing (TGRS), Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), ISPRS Journal of Photogrammetry and Remote Sensing, Computer Methods and Programs in Biomedicine (CMPB), Pattern Recognition Letters, Computer Vision and Image Understanding (CVIU), SPIE Optical Engineering (OE), Remote Sensing MDPI.

10.1.4. Invited Talks

- E. Chouzenoux:
  Polish academy of science, Warsaw, Poland, November 2019
  University of Edinburgh, UK, October 2019
  AIP 2019, Grenoble, France, July 2019
  SPARS 2019, Toulouse, France, 4 July 2019
  Inria Rennes PANAMA, April 2019
  Imagine ENPC, Champs-sur-Marne, April 2019
  Univ. Paris-Dauphine, Paris, March 2019
  Imaging in Paris, IHP, Paris, February 2019

- J.C. Pesquet:
  BASP Frontiers Workshop, Villars-sur-Ollon, Suisse, February 2019
  Modern Maximal Monotone Operator Theory workshop, Vienna, February 2019
  Workshop on Operator Splitting Methods in Data Analysis, New York, US, March 2019
  Marseille Université, Juin 2019
  International Conference on Optimisation, Berlin, August 2019
  EUSIPCO, A Coruña, Spain, September 2019
  French-German-Swiss Conference on Optimization, Nice, September 2019
  Wavelet and Applications workshop, Brussels, November 2019
- H. Talbot:
  CentraleSupélec: March 2019, November 2019
  Institut Henri Poincaré, Paris, February 2019
- M. Vakalopoulou:
  ARTORG Center for Biomedical Engineering Research, University of Bern, March 2019
  European Society of Thoracic Imaging 2019, May 2019
  Department of Computer Science, Stony Brook University, June 2019
  Electrical & Computer Engineering, University of Houston, June 2019
  France is AI, October 2019
  Intelligence Artificielle et Santé, Faculté de Médicine Paris-Sud, November 2019
  Annual Symposium of Belgian Association for Neuro-Oncology, December 2019

10.1.5. Leadership within the Scientific Community

J.-C. Pesquet is senior member of the Institut Universitaire de France and a Fellow of IEEE.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master : E. Chouzenoux. Foundations of Distributed and Large Scale Computing, 26h, M.Sc. in Data Sciences and Business Analytics, 3rd year CentraleSupélec, MVA ENS Cachan, Master Optimization Paris Sud and ESSEC Business School, FR
- Master: E. Chouzenoux. Advanced Machine Learning, 18h, 3rd year CentraleSupélec, FR
- Master: F. Malliaros. Machine Learning, 27h, M.Sc. in Data Sciences and Business Analytics, CentraleSupélec and ESSEC Business School and M.Sc. in Artificial Intelligence, CentraleSupélec, FR.
- Master: F. Malliaros. Introduction to Machine Learning, 33h, 2^{nd} year course at CentraleSupélec, FR.
- Master: F. Malliaros. Mathematical Modeling of Propagation Phenomena – Propagation on Graphs, 15h, 1^{st} year course at CentraleSupélec, FR.
- Master: Oyallon, Edouard. Deep Learning, 24h, 3rd year CentraleSupélec, FR
- Master: Oyallon, Edouard. Reinforcement Learning, 24h, 3rd year CentraleSupélec, FR
- Master : J.C. Pesquet. Advanced course on Optimization, 33h, M1, CentraleSupélec, FR
- Master: J.C. Pesquet. Introduction to Optimization, 6h, MVA ENS Cachan, FR
- Master: Talbot, Hugues. Discrete Optimisation, 2nd year course, CentraleSupelec, 30h, EN
- Master: Talbot, Hugues. Big Data, Techniques and Platforms, M.Sc in Data Science and Business Analytics, CentraleSupelec and ESSEC Business School, 30h, EN
- Master: Talbot, Hugues. Optimisation for AI, M.Sc in AI, CentraleSupelec and ESSEC Business School, 30h, EN
- Master: M. Vakalopoulou. Introduction to Visual Computing, 25h, 3rd year CentraleSupélec, FR

- Master: M. Vakalopoulou. Deep Learning, 25h, 3rd year CentraleSupélec, FR
- Master: M. Vakalopoulou. Introduction to Machine Learning, 33h, 2nd year CentraleSupélec, FR

10.2.2. Lecturing activities

- F. Malliaros. Summer School Artificial Intelligence, July 1-12, 2019, CentraleSupélec.
- J.C. Pesquet. Erwin Schrödinger Institute for Mathematics and Physics in Vienna, 4-6 March 2019, Austria.
- M. Vakalopoulou. Summer School on Artificial Intelligence, July 1-12n 2019, CentraleSupélec.
- E. Chouzenoux. Summer School Sparsity for Physics, Signal and Learning, June 24-27, 2019, Inria Paris.

10.2.3. Supervision

- PhD (defended) : Huu Dien Khue Le, Algorithme des directions alternées non convexe pour graphes: inférence et apprentissage, 2016-2019, supervised by N. Paragios.
- PhD (defended) : Riza Alp Guler, Apprentissage de Correspondances Image-Surface, 2016-2019, supervised by I. Kokkinos and N. Paragios.
- PhD (defended) : Marie-Caroline Corbineau, Fast online optimization algorithms for machine learning and medical imaging, 2016-2019, supervised by E. Chouzenoux and J.-C. Pesquet.
- PhD (defended) : Loubna El Gueddari, Parallel proximal algorithms for compressed sensing MRI reconstruction - Applications to ultra-high magnetic field imaging, 2016-2019, supervised by J.-C. Pesquet and Ph. Ciuciu (Inria PARIETAL).
- PhD (in progress) : Abdulkadir Çelikkanat, Representation learning methods on graphs, 2017-2020, supervised by F. Malliaros and N. Paragios.
- PhD (in progress): Yunshi Huang, Majorization-Minimization approaches for large scale problems in image processing, 2018-2021, supervised by E. Chouzenoux and V. Elvira (Univ. Edinburgh).
- PhD (in progress) : Samy Ammari, Imagerie médicale computationnelle en neuro oncologie, 2019-2022, supervised by C. Balleyguier (Institut Gustave Roussy) and E. Chouzenoux.
- PhD (in progress) : Georgios Panagopoulos, Influence maximization in social networks, 2018-2021, supervised by F. Malliaros and M. Vazirgiannis (École Polytechnique).
- PhD (in progress): Surabhi Jagtap, Graph-based learning from multi-omics data, 2019-2022, supervised by F. Malliaros, J.-C. Pesquet, and L. Duval (IFP Energies Nouvelles).
- PhD (in progress): Kavya Gupta, Neural network solutions for safety of complex systems, 2019-2022, supervised by F. Malliaros, J.-C. Pesquet and F. Kaakai (Thales Group).
- PhD (in progress): Maria Papadomanolaki, Change Detection from Multitemporal High Resolution Data with Deep Learning, 2017-2021, supervised by M. Vakalopoulou and with K. Karantzalos.
- PhD (in progress): Mihir Sahasrabudhe, Unsupervised ans Weakly Supervised Deep Learning Methods for Computer Vision and Medical Imaging, 2016-2020, supervised by N. Paragios.
- PhD (in progress): Théo Estienne, Improving anticancer therapies efficacy through Machine Learning on Medical Imaging & Genomic Data, 2017-2020, supervised by M. Vakalopoulou and N. Paragios.
- PhD (in progress): Enzo Battistella, Development of novel imaging approaches for tumour phenotype assessment by noninvasive imaging 2017-2020, supervised by M. Vakalopoulou and N. Paragios.
- PhD (in progress): Roger Sun, Deep learning and computer vision approaches on medical imaging and genomic data to improve the prediction of anticancer therapies' efficacy, 2017-2020, supervised by M. Vakalopoulou and N. Paragios.

- PhD (in progress): Ana Neacsu, Méthodes d'apprentissage profond inspirées d'algorithmes de traitement du signal, 2019-2022, supervised by J.-C. Pesquet and C. Burileanu (Politehnica Bucarest).
- PhD (in progress): Sagar Verma, Modélisation, contrôle et supervision de moteurs électriques par réseaux de neurones profonds, 2019-2022, supervised by M. Castella and J.-C. Pesquet.
- PhD (in progress): Maïssa Sghaier, clinical Task-Based Reconstruction in tomosynthesis, 2017-2020, supervised by J.-C. Pesquet and S. Muller (GE Healthcare).
- PhD (in progress): Marion Savanier, Reconstruction 3D interventionnelle, 2019-2022, supervised by E. Chouzenoux and C. Riddell (GE Healthcare).
- PhD (in progress): Arthur Marmin, Rational models optimized exactly for chemical processes improvement, 2017-2020, supervised by M. Castella (Telecom Paristech) and J.-C. Pesquet.
- PhD (in progress): Sylvain Lempereur. Analyse quantitative de la morphologie de poissons aux stades larvaire et juvénile. 2017-2020. Supervised by Hugues Talbot and Jean-Stéphane Joly (CNRS).
- PhD (in progress): Daniel Antunes: Contraintes géométriques et approches variationnelles pour l'analyse d'image. 2016-2019. Supervised by Hugues Talbot and Jacques-Olivier Lachaud (U. Savoie-Mont Blanc)
- PhD (in progress): Marie-Charlotte Poilpre: Méthode de comparaison faciale morphologique, adaptée aux expertise judiciaires, basée sur la modélisation 3D. 2017-2020. Supervised by Hugues Talbot and Vincent Nozick (U. Paris-Est)
- PhD (in progress): Thank Xuan Nguyen. Détection et étude morphologique des sources extra-galactiques par analyse variationnelle. 2018-2021. Supervised by Hugues Talbot and Laurent Najman (ESIEE)
- PhD (in progress): Marvin Lerousseau. Apprentissage statistique en imagerie médicale et en génomique pour prédire l'efficacité des thérapies anti-tumorales. 2018-2021. Supervised by Nikos Paragios (Therapanacea), Eric Deutch (IGR) and Hugues Talbot.
- PhD (in progress): Mario Viti. Low-dose assessment of coronal vessel health on CT. 2019-2022. Supervised by Hugues Talbot.

10.2.4. Juries

The faculty members of the team participated to numerous PhD Thesis Committees, HDR Committees and served as Grant Reviewers. Moreover, they serve regularly as a jury Member to Final Engineering Internship and the Research Innovation Project for students of CentraleSupélec, FR.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] M.-C. CORBINEAU.*Proximal and Interior Point Optimization Strategies in Image Recovery*, Université Paris-Saclay, Centrale Supélec, December 2019, https://hal.archives-ouvertes.fr/tel-02428404

[2] D. K. LÊ-HUU.*Nonconvex Alternating Direction Optimization for Graphs: Inference and Learning*, CentraleSupélec, Université Paris-Saclay, February 2019, https://hal.archives-ouvertes.fr/tel-02093465

Articles in International Peer-Reviewed Journal

[3] Ö. D. AKYILDIZ, E. CHOUZENOUX, V. ELVIRA, J. MÍGUEZ.*A probabilistic incremental proximal gradient method*, in "IEEE Signal Processing Letters", July 2019, vol. 26, n^{o} 8, p. 1257-1261, https://arxiv.org/abs/1812.01655 - 5 pages [*DOI :* 10.1109/LSP.2019.2926926], https://hal.archives-ouvertes.fr/hal-01946642

[4] A. BENFENATI, E. CHOUZENOUX, J.-C. PESQUET.*Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation*, in "Signal Processing", April 2020, vol. 169 [*DOI :* 10.1016/J.SIGPRO.2019.107417], https://hal.archives-ouvertes.fr/hal-02422403

[5] C. BERTOCCHI, E. CHOUZENOUX, M.-C. CORBINEAU, J.-C. PESQUET, M. PRATO.*Deep Unfolding of a Proximal Interior Point Method for Image Restoration*, in "Inverse Problems", 2019, https://arxiv.org/abs/1812.04276 , forthcoming [*DOI :* 10.1088/1361-6420/AB460A], https://hal.archives-ouvertes.fr/hal-01943475

[6] L. BRICEÑO-ARIAS, G. CHIERCHIA, E. CHOUZENOUX, J.-C. PESQUET.*A Random Block-Coordinate Douglas-Rachford Splitting Method with Low Computational Complexity for Binary Logistic Regression*, in "Computational Optimization and Applications", April 2019, vol. 72, n^{o} 3, p. 707-726, https://arxiv.org/abs/1712.09131 [*DOI :* 10.1007/S10589-019-00060-6], https://hal.archives-ouvertes.fr/hal-01672507

[7] M. CASTELLA, J.-C. PESQUET, A. MARMIN.*Rational optimization for nonlinear reconstruction with approximate $\ell 0$ penalization*, in "IEEE Transactions on Signal Processing", March 2019, vol. 67, n^{o} 6, p. 1407-1417 [*DOI :* 10.1109/TSP.2018.2890065], https://hal.archives-ouvertes.fr/hal-01852289

[8] G. CHASSAGNON, C. MARTIN, R. MARINI, M. VAKALOPOULOU, A. RÉGENT, L. MOUTHON, N. PARAGIOS, M.-P. REVEL.*Use of Elastic Registration in Pulmonary MRI for the Assessment of Pulmonary Fibrosis in Patients with Systemic Sclerosis*, in "Radiology", May 2019, vol. 291, n^{o} 2, p. 487-492 [*DOI :* 10.1148/RADIOL.2019182099], https://hal.inria.fr/hal-02422529

[9] G. CHASSAGNON, M. VAKALOPOULOU, N. PARAGIOS, M.-P. REVEL.*Artificial intelligence applications for thoracic imaging*, in "European Journal of Radiology", February 2020, vol. 123, 108774 [*DOI :* 10.1016/J.EJRAD.2019.108774], https://hal.inria.fr/hal-02422501

[10] E. CHOUZENOUX, M.-C. CORBINEAU, J.-C. PESQUET.*A Proximal Interior Point Algorithm with Applications to Image Processing*, in "Journal of Mathematical Imaging and Vision", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02120005

[11] E. CHOUZENOUX, H. GÉRARD, J.-C. PESQUET.*General risk measures for robust machine learning*, in "Foundations of Data Science", September 2019, vol. 1, n^{o} 3, p. 249-269, https://arxiv.org/abs/1904.11707 , https://hal.archives-ouvertes.fr/hal-02109418

[12] E. CHOUZENOUX, T. TSZ-KIT LAU, C. LEFORT, J.-C. PESQUET.*Optimal Multivariate Gaussian Fitting with Applications to PSF Modeling in Two-Photon Microscopy Imaging*, in "Journal of Mathematical Imaging and Vision", September 2019, vol. 61, n^{o} 7, p. 1037-1050 [*DOI :* 10.1007/S10851-019-00884-1], https://hal.archives-ouvertes.fr/hal-01985663

[13] P. L. COMBETTES, J.-C. PESQUET.*Deep Neural Network Structures Solving Variational Inequalities **, in "Set-Valued and Variational Analysis", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02425025

[14] P. L. COMBETTES, J.-C. PESQUET.*Stochastic Quasi-Fejér Block-Coordinate Fixed Point Iterations With Random Sweeping II: Mean-Square and Linear Convergence*, in "Mathematical Programming B", March 2019, vol. 174, n^{o} 1-2, p. 433–451, https://hal.archives-ouvertes.fr/hal-01964582

[15] M.-C. CORBINEAU, D. KOUAMÉ, E. CHOUZENOUX, J.-Y. TOURNERET, J.-C. PESQUET.*Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images - Extended Version*, in "IEEE Signal Processing Letters", August 2019, vol. 26, n^{o} 10, p. 1456–1460, https://hal.archives-ouvertes.fr/hal-02073283

[16] D. GENEST, É. PUYBAREAU, M. LÉONARD, J. COUSTY, N. DE CROZÉ, H. TALBOT.*High throughput automated detection of axial malformations in Medaka fish embryo*, in "Computers in Biology and Medicine", February 2019, vol. 105, p. 157-168 [*DOI :* 10.1016/J.COMPBIOMED.2018.12.016], https://hal.archives-ouvertes.fr/hal-01959606

[17] E. GROSSIORD, B. NAEGEL, H. TALBOT, L. NAJMAN, N. PASSAT.*Shape-based analysis on component-graphs for multivalued image processing*, in "Mathematical Morphology - Theory and Applications", 2019, vol. 3, n^{o} 1, p. 45-70 [*DOI :* 10.1515/MATHM-2019-0003], https://hal.univ-reims.fr/hal-01695384

[18] E. GROSSIORD, N. PASSAT, H. TALBOT, B. NAEGEL, S. KANOUN, I. TAL, P. TERVÉ, S. KEN, O. CASASNOVAS, M. MEIGNAN, L. NAJMAN.*Shaping for PET image analysis*, in "Pattern Recognition Letters", 2020, forthcoming, https://hal.archives-ouvertes.fr/hal-02155801

[19] C. JAQUET, L. NAJMAN, H. TALBOT, L. GRADY, M. SCHAAP, B. SPAIN, H. J. KIM, I. VIGNON-CLEMENTEL, C. A. TAYLOR.*Generation of patient-specific cardiac vascular networks: a hybrid image-based and synthetic geometric model*, in "IEEE Transactions on Biomedical Engineering", April 2019, vol. 66, n^{o} 4, p. 946-955 [*DOI :* 10.1109/TBME.2018.2865667], https://hal.archives-ouvertes.fr/hal-01869264

[20] F. MALLIAROS, C. GIATSIDIS, A. N. PAPADOPOULOS, M. VAZIRGIANNIS.*The Core Decomposition of Networks: Theory, Algorithms and Applications*, in "The VLDB Journal", 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-01986309

[21] A. MEKKI, L. DERCLE, P. LICHTENSTEIN, G. NASSER, A. MARABELLE, S. CHAMPIAT, E. CHOUZENOUX, C. BALLEYGUIER, S. AMMARI.*Machine learning defined diagnostic criteria for differentiating pituitary metastasis from autoimmune hypophysitis in patients undergoing immune checkpoint blockade therapy*, in "European Journal of Cancer", September 2019, vol. 119, p. 44-56 [*DOI :* 10.1016/J.EJCA.2019.06.020], https://hal.archives-ouvertes.fr/hal-02269518

[22] O. MERVEILLE, B. NAEGEL, H. TALBOT, N. PASSAT.*nD variational restoration of curvilinear structures with prior-based directional regularization*, in "IEEE Transactions on Image Processing", 2019, vol. 28, n^{o} 8, p. 3848-3859 [*DOI :* 10.1109/TIP.2019.2901706], https://hal.archives-ouvertes.fr/hal-01832636

[23] A. MONGIA, N. JHAMB, E. CHOUZENOUX, A. MAJUMDAR.*Deep latent factor model for collaborative filtering*, in "Signal Processing", 2020, vol. 169, 107366 [*DOI :* 10.1016/J.SIGPRO.2019.107366], https://hal.archives-ouvertes.fr/hal-02373934

[24] M. PAPADOMANOLAKI, M. VAKALOPOULOU, K. KARANTZALOS.*A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks*, in "Remote Sensing", March 2019, vol. 11, n^{o} 6, 684 [*DOI :* 10.3390/RS11060684], https://hal.inria.fr/hal-02078539

[25] J. ROBIC, B. PERRET, A. NKENGNE, M. COUPRIE, H. TALBOT.*Three-dimensional conditional random field for the dermal–epidermal junction segmentation*, in "Journal of Medical Imaging", April 2019, vol. 6, n^{o} 02, 1 [*DOI :* 10.1117/1.JMI.6.2.024003], https://hal.archives-ouvertes.fr/hal-02155490

[26] M. VAKALOPOULOU, G. CHASSAGNON, N. PARAGIOS, M.-P. REVEL.*Deep learning: definition and perspectives for thoracic imaging*, in "European Radiology", December 2019 [*DOI :* 10.1007/s00330-019-06564-3], https://hal.inria.fr/hal-02422513

Invited Conferences

[27] E. CHOUZENOUX, V. ELVIRA.*Adaptive importance sampling with scaled Langevin proposal adaptatioń*, in "17th International Conference on Computer Aided Systems Theory (EUROCAST 2019)", Las Palmas de Gran Canaria, Spain, February 2019, https://hal.archives-ouvertes.fr/hal-02314409

[28] A. MARMIN, M. CASTELLA, J.-C. PESQUET.*Detecting the rank of a symmetric tensor*, in "EUSIPCO 2019 : 27th European Signal Processing Conference", La Corogne, Spain, 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, 2019, p. 1-5 [*DOI :* 10.23919/EUSIPCO.2019.8902781], https://hal.archives-ouvertes.fr/hal-02284991

[29] Y. MARNISSI, E. CHOUZENOUX, A. BENAZZA-BENYAHIA, J.-C. PESQUET.*MM Adapted MH Methods*, in "BASP 2019 - International Biomedical and Astronomical Signal Processing Frontiers workshop", Villars sur Ollon, Switzerland, February 2019, https://hal.archives-ouvertes.fr/hal-02314412

[30] M. TERRIS, E. CHOUZENOUX.*Stochastic MM Subspace Algorithms*, in "BASP 2019 - International Biomedical and Astronomical Signal Processing Frontiers workshop", Villars sur Ollon, Switzerland, February 2019, https://hal.archives-ouvertes.fr/hal-02314411

International Conferences with Proceedings

[31] D. ANTUNES, J.-O. LACHAUD, H. TALBOT.*Digital Curvature Evolution Model for Image Segmentation*, in "International Conference on Discrete Geometry for Computer Imagery", Noisy-le-Grand, France, Lecture Notes in Computer Science, Springer, February 2019, vol. 11414, p. 15-26 [*DOI :* 10.1007/978-3-030-14085-4_2], https://hal.archives-ouvertes.fr/hal-02426946

[32] E. BATTISTELLA, M. VAKALOPOULOU, T. ESTIENNE, M. LEROUSSEAU, R. SUN, C. ROBERT, N. PARAGIOS, E. DEUTSCH.*Gene Expression High-Dimensional Clustering towards a Novel, Robust, Clinically Relevant and Highly Compact Cancer Signature*, in "IWBBIO 2019 - 7th International Work-Conference on Bioinformatics and Biomedical Engineering", Granada, Spain, 2019, https://hal.archives-ouvertes.fr/hal-02076104

[33] E. BELILOVSKY, M. EICKENBERG, E. OYALLON.*Greedy Layerwise Learning Can Scale to ImageNet*, in "ICML 2019 - 36th International Conference on Machine Learning", Long Beach, CA, United States, June 2019, https://arxiv.org/abs/1812.11446 , https://hal.inria.fr/hal-02119398

[34] A. BENAMIRA, B. DEVILLERS, E. LESOT, A. K. RAY, M. SAADI, F. MALLIAROS.*Semi-Supervised Learning and Graph Neural Networks for Fake News Detection*, in "ASONAM 2019 - IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining", Vancouver, Canada, August 2019, https://hal.archives-ouvertes.fr/hal-02334445

[35] A. CHERNI, E. CHOUZENOUX, L. DUVAL, J.-C. PESQUET.*A Novel Smoothed Norm Ratio for Sparse Signal Restoration Application to Mass Spectrometry*, in "SPARS (Signal Processing with Adaptive Sparse Structured Representations)", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02179379

[36] M.-C. CORBINEAU, C. BERTOCCHI, E. CHOUZENOUX, M. PRATO, J.-C. PESQUET.*Learned Image Deblurring by Unfolding a Proximal Interior Point Algorithm*, in "ICIP 2019 - 26th IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019 [*DOI :* 10.1109/ICIP.2019.8803438], https://hal.archives-ouvertes.fr/hal-02303511

[37] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*Calibrationless oscar-based image reconstruction in compressed sensing parallel MRI*, in "ISBI 2019 - IEEE International Symposium on Biomedical Imaging", Venise, Italy, April 2019, https://hal.inria.fr/hal-02101262

[38] V. ELVIRA, E. CHOUZENOUX.*Langevin-based Strategy for Efficient Proposal Adaptation in Population Monte Carlo*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2019)", Brighton, United Kingdom, May 2019 [*DOI :* 10.1109/ICASSP.2019.8682284], https://hal.archives-ouvertes.fr/hal-02431677

[39] D. GENEST, M. LÉONARD, J. COUSTY, N. DE CROZÉ, H. TALBOT.*Atlas-based automated detection of swim bladder in Medaka embryo*, in "International Symposium on Mathematical Morphology (ISMM)", Saarbrücken, Germany, Lecture Notes in Computer Science, Springer, May 2019, vol. 11564, p. 496-507, https://arxiv.org/abs/1902.06130 [*DOI :* 10.1007/978-3-030-20867-7_38], https://hal.archives-ouvertes.fr/hal-02019658

[40] L. E. GUEDDARI, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET, P. CIUCIU.*Online MR image reconstruction for compressed sensing acquisition in T2* imaging*, in "SPIE Conference - Wavelets and Sparsity XVIII", San Diego, United States, August 2019, https://hal.inria.fr/hal-02265538

[41] Y. HUANG, E. CHOUZENOUX, V. ELVIRA.*Particle Filtering for Online Space-Varying Blur Identification*, in "IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing", Le Gosier, France, Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, December 2019, https://hal.archives-ouvertes.fr/hal-02406970

[42] A. MARMIN, M. CASTELLA, J.-C. PESQUET.*Sparse signal reconstruction with a sign oracle*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations - Workshop", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02196881

[43] Y. MARNISSI, D. ABBOUD, E. CHOUZENOUX, J.-C. PESQUET, M. EL-BADAOUI, A. BENAZZA-BENYAHIA.*A Data Augmentation Approach for Sampling Gaussian Models in High Dimension*, in "EUSIPCO 2019 - 27th European Signal Processing Conference", La Corogne, Spain, Proceedings of the 27th European Signal Processing Conference (EUSIPCO 2019), September 2019, https://hal.archives-ouvertes.fr/hal-02314418

[44] A. MONGIA, V. JAIN, E. CHOUZENOUX, A. MAJUMDAR.*Deep Latent Factor Model for Predicting Drug Target Interactions*, in "IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, May 2019, https://hal.archives-ouvertes.fr/hal-02406989

[45] M. PAPADOMANOLAKI, K. KARANTZALOS, M. VAKALOPOULOU.*A multi-task deep learning framework coupling semantic segmentation and image reconstruction for very high resolution imagery*, in "IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium", Yokohama, Japan, July 2019, https://hal.inria.fr/hal-02266085

[46] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, K. Karantzalos.*Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data*, in "IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium", Yokohama, Japan, July 2019, https://hal.inria.fr/hal-02266094

[47] É. Puybareau, E. Carlinet, A. Benfenati, H. Talbot.*Spherical Fluorescent Particle Segmentation and Tracking in 3D Confocal Microscopy*, in "International Symposium on Mathematical Morphology, ISMM 2019", Saarbrücken, Germany, Lecture Notes in Computer Science, Springer, May 2019, vol. 11564, p. 520-531 [*DOI :* 10.1007/978-3-030-20867-7_40], https://hal.archives-ouvertes.fr/hal-02426948

[48] J. Robic, B. Perret, A. Nkengne, M. Couprie, H. Talbot.*Self-dual pattern spectra for characterising the dermal-epidermal junction in 3D reflectance confocal microscopy imaging*, in "International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing", Saarbrücken, Germany, Lecture Notes in Computer Science, Springer, May 2019, vol. 11564, p. 508-519 [*DOI :* 10.1007/978-3-030-20867-7_39], https://hal.archives-ouvertes.fr/hal-02169702

[49] M. Sahasrabudhe, Z. Shu, E. Bartrum, R. A. Güler, D. Samaras, I. Kokkinos.*Lifting AutoEncoders: Unsupervised Learning of a Fully-Disentangled 3D Morphable Model Using Deep Non-Rigid Structure From Motion*, in "The IEEE International Conference on Computer Vision (ICCV) Workshops", Seoul, South Korea, The IEEE International Conference on Computer Vision (ICCV) Workshops, November 2019, https://hal.archives-ouvertes.fr/hal-02422596

[50] M. Sghaier, E. Chouzenoux, G. Palma, J.-C. Pesquet, S. Muller.*A New Approach for Microcalcification Enhancement in Digital Breast Tomosynthesis Reconstruction*, in "ISBI 2019 - IEEE International Symposium on Biomedical Imaging", Venise, Italy, Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI 2019), April 2019 [*DOI :* 10.1109/ISBI.2019.8759534], https://hal.archives-ouvertes.fr/hal-02314420

[51] A. J.-P. Tixier, M. E. G. Rossi, F. Malliaros, J. Read, M. Vazirgiannis.*Perturb and Combine to Identify Influential Spreaders in Real-World Networks*, in "IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining", Vancouver, Canada, August 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-01958973

[52] M. Vakalopoulou, S. Christodoulidis, M. Sahasrabudhe, S. Mougiakakou, N. Paragios.*Image Registration of Satellite Imagery with Deep Convolutional Neural Networks*, in "IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium", Yokohama, France, IEEE, July 2019, p. 4939-4942 [*DOI :* 10.1109/IGARSS.2019.8898220], https://hal.inria.fr/hal-02422555

[53] A. Çelikkanat, F. Malliaros.*Kernel Node Embeddings*, in "IEEE Global Conference on Signal and Information Processing (GlobalSIP)", Ottawa, Canada, November 2019, https://hal.archives-ouvertes.fr/hal-02423629

[54] A. Çelikkanat, F. Malliaros.*Learning Node Embeddings with Exponential Family Distributions*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems - Workshop on Graph Representation Learning", Vancouver, Canada, December 2019, https://hal.archives-ouvertes.fr/hal-02336000

Conferences without Proceedings

[55] A. CHERNI, E. CHOUZENOUX, L. DUVAL, J.-C. PESQUET.*Forme lissée de rapports de normes lp/lq (SPOQ) pour la reconstruction des signaux avec pénalisation parcimonieuse*, in "GRETSI 2019", Lille, France, August 2019, https://hal.archives-ouvertes.fr/hal-02179373

[56] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*Online compressed sensing MR image reconstruction for high resolution T2* imaging*, in "ISMRM 27th Annual Meeting and Exhibition", Montréal, Canada, May 2019, https://hal.archives-ouvertes.fr/hal-02314904

[57] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*OSCAR-based reconstruction for compressed sensing and parallel MR imaging*, in "ISMRM 27th Annual Meeting and Exhibition", Montréal, Canada, May 2019, https://hal.archives-ouvertes.fr/hal-02314911

[58] T. ESTIENNE, M. VAKALOPOULOU, S. CHRISTODOULIDIS, E. BATTISTELLA, M. LEROUSSEAU, A. CARRE, G. KLAUSNER, R. SUN, C. ROBERT, S. MOUGIAKAKOU, N. PARAGIOS, E. DEUTSCH.*U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets*, in "MICCAI 2019: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019", Shenzhen, China, MICCAI, October 2019, p. 310-319 [*DOI :* 10.1007/978-3-030-32248-9_35], https://hal.archives-ouvertes.fr/hal-02365899

[59] C. LEFORT, E. CHOUZENOUX, J.-C. PESQUET.*PLUMEE 2019 : Reconstruction numérique d'image en microscopie multiphotonique*, in "6ème Colloque francophone PLUridisciplinaire sur les Matériaux, l'Environnement et l'Electronique (PLUMEE 2019)", Limoges, France, April 2019, https://hal.archives-ouvertes.fr/hal-02426765

[60] S. VERMA, R. VERMA, P. B. SUJIT.*MAPEL: Multi-Agent Pursuer-Evader Learning using Situation Report*, in "2019 International Joint Conference on Neural Networks (IJCNN)", Budapest, France, IEEE, July 2019, p. 1-8 [*DOI :* 10.1109/IJCNN.2019.8852457], https://hal.archives-ouvertes.fr/hal-02426661

Other Publications

[61] A. BENFENATI, F. BONACCI, T. BOUROUINA, H. TALBOT.*Efficient position estimation of 3D fluorescent spherical beads in confocal microscopy via Poisson denoising*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02150316

[62] L. EL GUEDDARI, E. CHOUZENOUX, A. VIGNAUD, P. CIUCIU.*Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02292372

[63] B. KAS, H. TALBOT, R. FERRARA, C. RICHARD, B. BESSE, L. MEZQUITA, N. LASSAU, C. CARAMELLA.*Hyperprogressive disease during immunotherapy: an attempt to clarify different definitions*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02429762

Project-Team PARIETAL

Modelling brain structure, function and variability based on high-field MRI data.

IN COLLABORATION WITH: CEA Neurospin

IN PARTNERSHIP WITH:
Centre CEA-Saclay

RESEARCH CENTER
Saclay - Île-de-France

THEME
Computational Neuroscience and Medicine

Table of contents

Project-Team PARIETAL

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

A3.3. - Data and knowledge analysis
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A3.4.8. - Deep learning
A5.3.2. - Sparse modeling and image representation
A5.3.3. - Pattern recognition
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.9.6. - Optimization tools
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

B1.2. - Neuroscience and cognitive science
B1.2.1. - Understanding and simulation of the brain and the nervous system
B1.2.2. - Cognitive science
B2.2.6. - Neurodegenerative diseases
B2.6.1. - Brain imaging

1. Team, Visitors, External Collaborators

Research Scientists

Bertrand Thirion [Team leader, Inria, Senior Researcher, HDR]
Philippe Ciuciu [CEA, Senior Researcher, HDR]
Denis Alexander Engemann [Inria, Starting Research Position]
Alexandre Gramfort [Inria, Senior Researcher, HDR]
Thomas Moreau [Inria, Researcher, from Oct 2019]
Gaël Varoquaux [Inria, Researcher, HDR]
Demian Wassermann [Inria, Researcher]

Faculty Member

Anna Kazeykina [Univ Paris-Saclay, Associate Professor, from Sep 2019]

External Collaborators

Danilo Bzdok [Univ. Aachen]
Luigi Gresele [MPI, from Oct 2019]
Joseph Salmon [Institut Telecom ex GET Groupe des Écoles des Télécommunications, HDR]
Darya Chyzhyk [Univ. San Sebastian]
Meyer Scetbon [ENS Paris Saclay]
Elvis Dohmatob [Criteo]
Arthur Mensch [Ecole Normale Supérieure]

Technical Staff

Alexandre Abadie [Inria, Engineer]
Nicolas Chartier [CEA, Engineer, until Aug 2019]
Kshitij Chawla [Inria, Engineer]
L Emir Omar Chehab [Inria, Engineer, from Dec 2019]
Guillaume Daval Frerot [CEA, Engineer, from Nov 2019]
Jérémie Du Boisberranger [Inria, Engineer]
Merlin Dumeur [École Normale Supérieure de Lyon, Engineer, from Nov 2019]
Guillaume Favelier [Inria, Engineer]
Chaithya Giliyar Radhkrishna [CEA, Engineer, from Jul 2019]
Pierre Glaser [Inria, Engineer, until Sep 2019]
Ana Luisa Grilo Pinho [Inria, Engineer]
Olivier Grisel [Inria, Engineer, until Jan 2019]
Guillaume Lemaitre [Inria, Engineer, until Mar 2019]
Jiaping Liu [Inria, Engineer, from Jul 2019]
Joan Massich Vall [Inria, Engineer, until Oct 2019]
Romuald Menuet [Inria, Engineer, until May 2019]
Thomas Schmitt [Inria, Engineer, from Aug 2019]
Swetha Shankar [Inria, Engineer, from Oct 2019]
Maria Telenczuk [Inria, Engineer, from Feb 2019]
Juan Jesus Torre Tresols [Inria, Engineer]
Joris Van Den Bossche [Inria, Engineer, until Mar 2019]

PhD Students

Pierre Ablin [Inria, PhD Student]
Thomas Bazeille [Inria, PhD Student]
Quentin Bertrand [Inria, PhD Student]
Patricio Cerda Reyes [Inria, PhD Student, until Oct 2019]
Hamza Cherkaoui [CEA, PhD Student]
Jerome-Alexis Chevalier [Inria, PhD Student]
Alexis Cvetkov Iliev [Inria, PhD Student, from Oct 2019]
Kamalaker Reddy Dadi [Inria, PhD Student]
Jérome Dockès [Inria, PhD Student]
Loubna El Gueddari [Ministère de l'Enseignement Supérieur et de la Recherche, PhD Student]
Valentin Iovene [Inria, PhD Student]
Hubert Jacob Banville [Interaxon, PhD Student]
Maeliss Jallais [Inria, PhD Student]
Hicham Janati [Inria, PhD Student]
Antonia Machlouzarides Shalit [Inria, PhD Student]
Lucas Martin [Inria, PhD Student, from Oct 2019]
Mathurin Massias [Inria, PhD Student]
Tuan Binh Nguyen [Inria, PhD Student]
Nicolas Prost [École polytechnique, PhD Student]

Zaccharie Ramzi [CEA, PhD Student, from Feb 2019]
Hugo Richard [Univ Paris-Saclay, PhD Student]
David Sabbagh [INSERM, PhD Student]
Gaston Zanitti [Inria, PhD Student, from Jun 2019]

Post-Doctoral Fellow
Marine Le Morvan [CNRS, Post-Doctoral Fellow, from May 2019]

Visiting Scientist
Huchuan Xia [Univ. Pennsylvania, from Jun 2019 until Sep 2019]

Administrative Assistant
Corinne Petitot [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

The Parietal team focuses on mathematical methods for modeling and statistical inference based on neuroimaging data, with a particular interest in machine learning techniques and applications of human functional imaging. This general theme splits into four research axes:

- Modeling for neuroimaging population studies,
- Encoding and decoding models for cognitive imaging,
- Statistical and machine learning methods for large-scale data,
- Compressed-sensing for MRI.

Parietal is also strongly involved in open-source software development in scientific Python (machine learning) and for neuroimaging applications.

3. Research Program

3.1. Inverse problems in Neuroimaging

Many problems in neuroimaging can be framed as forward and inverse problems. For instance, brain population imaging is concerned with the *inverse problem* that consists in predicting individual information (behavior, phenotype) from neuroimaging data, while the corresponding *forward problem* boils down to explaining neuroimaging data with the behavioral variables. Solving these problems entails the definition of two terms: a loss that quantifies the goodness of fit of the solution (does the model explain the data well enough?), and a regularization scheme that represents a prior on the expected solution of the problem. These priors can be used to enforce some properties on the solutions, such as sparsity, smoothness or being piece-wise constant.

Let us detail the model used in typical inverse problem: Let $\mathbf{X}$ be a neuroimaging dataset as an $(n_{subjects}, n_{voxels})$ matrix, where $n_{subjects}$ and n_{voxels} are the number of subjects under study, and the image size respectively, $\mathbf{Y}$ a set of values that represent characteristics of interest in the observed population, written as $(n_{subjects}, n_{features})$ matrix, where $n_{features}$ is the number of characteristics that are tested, and $\mathbf{w}$ an array of shape $(n_{voxels}, n_{features})$ that represents a set of pattern-specific maps. In the first place, we may consider the columns $\mathbf{Y}_1, .., \mathbf{Y}_{n_{features}}$ of Y independently, yielding $n_{features}$ problems to be solved in parallel:

$$\mathbf{Y}_i = \mathbf{X}\mathbf{w}_i + \epsilon_i, \forall i \in \{1, .., n_{features}\},$$

where the vector contains $\mathbf{w}_i$ is the i^{th} row of $\mathbf{w}$. As the problem is clearly ill-posed, it is naturally handled in a regularized regression framework:

$$\widehat{w}_i = \text{argmin}_{w_i} \|\mathbf{Y}_i - \mathbf{X}\mathbf{w}_i\|^2 + \Psi(\mathbf{w}_i), \tag{1}$$

where Ψ is an adequate penalization used to regularize the solution:

$$\Psi(\mathbf{w}; \lambda_1, \lambda_2, \eta_1, \eta_2) = \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2 + \eta_1 \|\nabla \mathbf{w}\|_{2,1} + \eta_2 \|\nabla \mathbf{w}\|_{2,2} \tag{2}$$

with λ_1, λ_2, η_1, $\eta_2 \geq 0$ (this formulation particularly highlights the fact that convex regularizers are norms or quasi-norms). In general, only one or two of these constraints is considered (hence is enforced with a non-zero coefficient):

- When $\lambda_1 > 0$ only (LASSO), and to some extent, when $\lambda_1, \lambda_2 > 0$ only (elastic net), the optimal solution $\mathbf{w}$ is (possibly very) sparse, but may not exhibit a proper image structure; it does not fit well with the intuitive concept of a brain map.
- Total Variation regularization (see Fig. 1) is obtained for ($\eta_1 > 0$ only), and typically yields a piece-wise constant solution. It can be associated with Lasso to enforce both sparsity and sparse variations.
- Smooth lasso is obtained with ($\eta_2 > 0$ and $\lambda_1 > 0$ only), and yields smooth, compactly supported spatial basis functions.

Note that, while the qualitative aspect of the solutions are very different, the predictive power of these models is often very close.

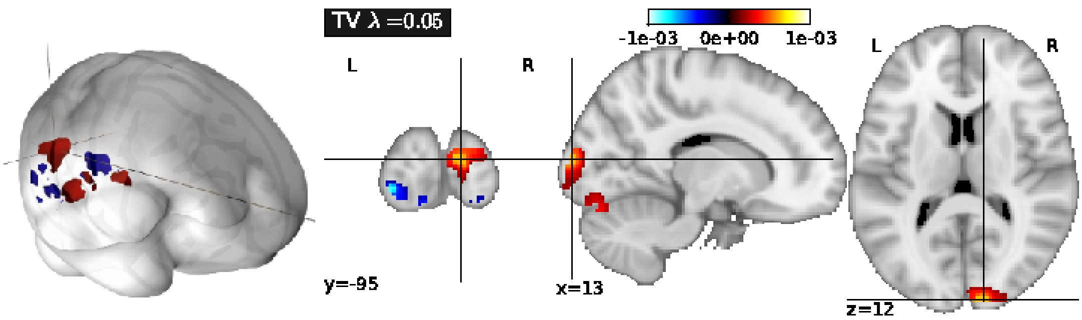


Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problem here is to predict the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the presentation of an image. Learning and test are performed across individuals. Unlike other approaches, Total Variation regularization yields a sparse and well-localized solution that also enjoys high predictive accuracy.

The performance of the predictive model can simply be evaluated as the amount of variance in $\mathbf{Y}_i$ fitted by the model, for each $i \in \{1, .., n_{features}\}$. This can be computed through cross-validation, by *learning*$\widehat{\mathbf{w}}_i$ on some part of the dataset, and then estimating $\|\mathbf{Y}_i - \mathbf{X}\widehat{w}_i\|^2$ using the remainder of the dataset.

This framework is easily extended by considering

- *Grouped penalization*, where the penalization explicitly includes a prior clustering of the features, i.e. voxel-related signals, into given groups. This amounts to enforcing structured priors on the solution.
- *Combined penalizations*, i.e. a mixture of simple and group-wise penalizations, that allow some variability to fit the data in different populations of subjects, while keeping some common constraints.
- *Logistic and hinge regression*, where a non-linearity is applied to the linear model so that it yields a probability of classification in a binary classification problem.
- *Robustness to between-subject variability* to avoid the learned model overly reflecting a few outlying particular observations of the training set. Note that noise and deviating assumptions can be present in both $\mathbf{Y}$ and $\mathbf{X}$
- *Multi-task learning*: if several target variables are thought to be related, it might be useful to constrain the estimated parameter vector $\mathbf{w}$ to have a shared support across all these variables.
  For instance, when one of the variables $\mathbf{Y}_i$ is not well fitted by the model, the estimation of other variables $\mathbf{Y}_j, j \neq i$ may provide constraints on the support of $\mathbf{w}_i$ and thus, improve the prediction of $\mathbf{Y}_i$.

$$\mathbf{Y} = \mathbf{X}\mathbf{w} + \epsilon, \tag{3}$$

then

$$\widehat{w} = \text{argmin}_{\mathbf{w}=(\mathbf{w}_i),i=1..n_f} \sum_{i=1}^{n_f} \|\mathbf{Y_i} - \mathbf{X}\mathbf{w_i}\|^2 + \lambda \sum_{j=1}^{n_{voxels}} \sqrt{\sum_{i=1}^{n_f} \mathbf{w_{i,j}^2}} \tag{4}$$

3.2. Multivariate decompositions

Multivariate decompositions provide a way to model complex data such as brain activation images: for instance, one might be interested in extracting an *atlas of brain regions* from a given dataset, such as regions exhibiting similar activity during a protocol, across multiple protocols, or even in the absence of protocol (during resting-state). These data can often be factorized into spatial-temporal components, and thus can be estimated through *regularized Principal Components Analysis* (PCA) algorithms, which share some common steps with regularized regression.
Let $\mathbf{X}$ be a neuroimaging dataset written as an $(n_{subjects}, n_{voxels})$ matrix, after proper centering; the model reads

$$\mathbf{X} = \mathbf{A}\mathbf{D} + \epsilon, \tag{5}$$

where $\mathbf{D}$ represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}), and $\mathbf{A}$ the associated subject-wise loadings. While traditional PCA and independent components analysis (ICA) are limited to reconstructing components $\mathbf{D}$ within the space spanned by the column of $\mathbf{X}$, it seems desirable to add some constraints on the rows of $\mathbf{D}$, that represent spatial maps, such as sparsity, and/or smoothness, as it makes the interpretation of these maps clearer in the context of neuroimaging. This yields the following estimation problem:

$$\min_{\mathbf{D},\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{D}\|^2 + \Psi(\mathbf{D}) \text{ s.t. } \|\mathbf{A}_i\| = 1 \; \forall i \in \{1..n_{features}\}, \tag{6}$$

where $(\mathbf{A}_i),\ i \in \{1..n_{features}\}$ represents the columns of $\mathbf{A}$. Ψ can be chosen such as in Eq. (2) in order to enforce smoothness and/or sparsity constraints.

The problem is not jointly convex in all the variables but each penalization given in Eq (2) yields a convex problem on $\mathbf{D}$ for $\mathbf{A}$ fixed, and conversely. This readily suggests an alternate optimization scheme, where $\mathbf{D}$ and $\mathbf{A}$ are estimated in turn, until convergence to a local optimum of the criterion. As in PCA, the extracted components can be ranked according to the amount of fitted variance. Importantly, also, estimated PCA models can be interpreted as a probabilistic model of the data, assuming a high-dimensional Gaussian distribution (probabilistic PCA).

Ultimately, the main limitations to these algorithms is the cost due to the memory requirements: holding datasets with large dimension and large number of samples (as in recent neuroimaging cohorts) leads to inefficient computation. To solve this issue, online methods are particularly attractive [1].

3.3. Covariance estimation

Another important estimation problem stems from the general issue of learning the relationship between sets of variables, in particular their covariance. Covariance learning is essential to model the dependence of these variables when they are used in a multivariate model, for instance to study potential interactions among them and with other variables. Covariance learning is necessary to model latent interactions in high-dimensional observation spaces, e.g. when considering multiple contrasts or functional connectivity data.
The difficulties are two-fold: on the one hand, there is a shortage of data to learn a good covariance model from an individual subject, and on the other hand, subject-to-subject variability poses a serious challenge to the use of multi-subject data. While the covariance structure may vary from population to population, or depending on the input data (activation versus spontaneous activity), assuming some shared structure across problems, such as their sparsity pattern, is important in order to obtain correct estimates from noisy data. Some of the most important models are:

- **Sparse Gaussian graphical models**, as they express meaningful conditional independence relationships between regions, and do improve conditioning/avoid overfit.
- **Decomposable models**, as they enjoy good computational properties and enable intuitive interpretations of the network structure. Whether they can faithfully or not represent brain networks is still an open question.
- **PCA-based regularization of covariance** which is powerful when modes of variation are more important than conditional independence relationships.

Adequate model selection procedures are necessary to achieve the right level of sparsity or regularization in covariance estimation; the natural evaluation metric here is the out-of-sample likelihood of the associated Gaussian model. Another essential remaining issue is to develop an adequate statistical framework to test differences between covariance models in different populations. To do so, we consider different means of parametrizing covariance distributions and how these parametrizations impact the test of statistical differences across individuals.

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain , such as: How is the function of cortical territories constrained by anatomy (connectivity) ? How to assess the specificity of brain regions ? How can one characterize reliably inter-subject differences ?

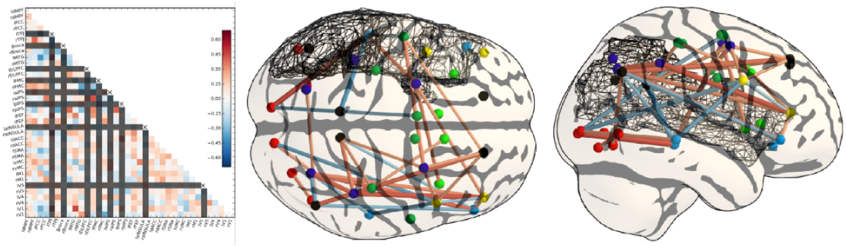


Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion volume outlined as a mesh) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical detection of the difference requires a sophisticated statistical framework for the comparison of graphical models.

4.1.2. Analysis of brain Connectivity

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. Modeling of brain processes (MEG)

While fMRI has been very useful in defining the function of regions at the mm scale, Magneto-encephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. Current challenges in human neuroimaging (acquisition+analysis)

Human neuroimaging targets two major goals: *i)* the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii)* the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial [0] resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.
- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements [0] and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [72], and

[0] and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

[0] The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [73]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as NeuroVault or OpenfMRI. A framework for integrating such datasets is however still missing.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- November 2019: Académie des sciences / Dassault System / Inria prize awarded to the developers of scikit-learn: Loic Estève, Alexandre Gramfort, Olivier Grisel, Bertrand Thirion and Gael Varoquaux.
- December 2019: Alexandre Gramfort, Bertrand Thirion and Gael Varoquaux are each awarded a *Chaire IA* following a national call.
- December 2019: Carole Lazarus, former PhD student supervised by Philippe Ciuciu got the Prix de la Chancellerie des Universités de Paris 2019 - Section Sciences.

6. New Software and Platforms

6.1. Mayavi

FUNCTIONAL DESCRIPTION: Mayavi is the most used scientific 3D visualization Python software. Mayavi can be used as a visualization tool, through interactive command line or as a library. It is distributed under Linux through Ubuntu, Debian, Fedora and Mandriva, as well as in PythonXY and EPD Python scientific distributions. Mayavi is used by several software platforms, such as PDE solvers (fipy, sfepy), molecule visualization tools and brain connectivity analysis tools (connectomeViewer).

- Contact: Gaël Varoquaux
- URL: http://mayavi.sourceforge.net/

6.2. Nilearn

NeuroImaging with scikit learn

KEYWORDS: Health - Neuroimaging - Medical imaging

FUNCTIONAL DESCRIPTION: NiLearn is the neuroimaging library that adapts the concepts and tools of scikit-learn to neuroimaging problems. As a pure Python library, it depends on scikit-learn and nibabel, the main Python library for neuroimaging I/O. It is an open-source project, available under BSD license. The two key components of NiLearn are i) the analysis of functional connectivity (spatial decompositions and covariance learning) and ii) the most common tools for multivariate pattern analysis. A great deal of efforts has been put on the efficiency of the procedures both in terms of memory cost and computation time.

- Participants: Alexandre Abraham, Alexandre Gramfort, Bertrand Thirion, Elvis Dohmatob, Fabian Pedregosa Izquierdo, Gaël Varoquaux, Loïc Estève, Michael Eickenberg and Virgile Fritsch
- Contact: Bertrand Thirion
- URL: http://nilearn.github.io/

6.3. Scikit-learn

KEYWORDS: Regession - Clustering - Learning - Classification - Medical imaging

SCIENTIFIC DESCRIPTION: Scikit-learn is a Python module integrating classic machine learning algorithms in the tightly-knit scientific Python world. It aims to provide simple and efficient solutions to learning problems, accessible to everybody and reusable in various contexts: machine-learning as a versatile tool for science and engineering.

FUNCTIONAL DESCRIPTION: Scikit-learn can be used as a middleware for prediction tasks. For example, many web startups adapt Scikitlearn to predict buying behavior of users, provide product recommendations, detect trends or abusive behavior (fraud, spam). Scikit-learn is used to extract the structure of complex data (text, images) and classify such data with techniques relevant to the state of the art.

Easy to use, efficient and accessible to non datascience experts, Scikit-learn is an increasingly popular machine learning library in Python. In a data exploration step, the user can enter a few lines on an interactive (but non-graphical) interface and immediately sees the results of his request. Scikitlearn is a prediction engine . Scikit-learn is developed in open source, and available under the BSD license.

- Participants: Alexandre Gramfort, Bertrand Thirion, Fabian Pedregosa Izquierdo, Gaël Varoquaux, Loïc Estève, Michael Eickenberg and Olivier Grisel
- Partners: CEA - Logilab - Nuxeo - Saint Gobain - Tinyclues - Telecom Paris
- Contact: Olivier Grisel
- URL: http://scikit-learn.org

6.4. MODL

Massive Online Dictionary Learning

KEYWORDS: Pattern discovery - Machine learning

FUNCTIONAL DESCRIPTION: Matrix factorization library, usable on very large datasets, with optional sparse and positive factors.

- Participants: Arthur Mensch, Gaël Varoquaux, Bertrand Thirion and Julien Mairal
- Contact: Arthur Mensch
- Publications: Subsampled online matrix factorization with convergence guarantees - hal-01431618v3
- URL: http://github.com/arthurmensch/modl

6.5. MNE

MNE-Python

KEYWORDS: Neurosciences - EEG - MEG - Signal processing - Machine learning

FUNCTIONAL DESCRIPTION: Open-source Python software for exploring, visualizing, and analyzing human neurophysiological data: MEG, EEG, sEEG, ECoG, and more.

RELEASE FUNCTIONAL DESCRIPTION: http://martinos.org/mne/stable/whats_new.html

- Partners: HARVARD Medical School - New York University - University of Washington - CEA - Aalto university - Telecom Paris - Boston University - UC Berkeley
- Contact: Alexandre Gramfort
- URL: http://martinos.org/mne/

6.6. Dmipy

Diffusion MRI Multi-Compartment Modeling and Microstructure Recovery Made Easy

KEYWORDS: Diffusion MRI - Multi-Compartment Modeling - Microstructure Recovery

FUNCTIONAL DESCRIPTION: Non-invasive estimation of brain microstructure features using diffusion MRI (dMRI) – known as Microstructure Imaging – has become an increasingly diverse and complicated field over the last decades. Multi-compartment (MC)-models, representing the measured diffusion signal as a linear combination of signal models of distinct tissue types, have been developed in many forms to estimate these features. However, a generalized implementation of MC-modeling as a whole, providing deeper insights in its capabilities, remains missing. To address this fact, we present Diffusion Microstructure Imaging in Python (Dmipy), an open-source toolbox implementing PGSE-based MC-modeling in its most general form. Dmipy allows on-the-fly implementation, signal modeling, and optimization of any user-defined MC-model, for any PGSE acquisition scheme. Dmipy follows a “building block”-based philosophy to Microstructure Imaging, meaning MC-models are modularly constructed to include any number and type of tissue models, allowing simultaneous representation of a tissue’s diffusivity, orientation, volume fractions, axon orientation dispersion, and axon diameter distribution. In particular, Dmipy is geared toward facilitating reproducible, reliable MC-modeling pipelines, often allowing the whole process from model construction to parameter map recovery in fewer than 10 lines of code. To demonstrate Dmipy’s ease of use and potential, we implement a wide range of well-known MC-models, including IVIM, AxCaliber, NODDI(x), Bingham-NODDI, the spherical mean-based SMT and MC-MDI, and spherical convolution-based single- and multi-tissue CSD. By allowing parameter cascading between MC-models, Dmipy also facilitates implementation of advanced approaches like CSD with voxel-varying kernels and single-shell 3-tissue CSD. By providing a well-tested, user-friendly toolbox that simplifies the interaction with the otherwise complicated field of dMRI-based Microstructure Imaging, Dmipy contributes to more reproducible, high-quality research.

- Authors: Rutger Fick, Demian Wassermann and Rachid Deriche
- Contact: Rachid Deriche

6.7. PySAP

Python Sparse data Analysis Package

KEYWORDS: Image reconstruction - Image compression

FUNCTIONAL DESCRIPTION: The PySAP (Python Sparse data Analysis Package, https://github.com/CEA-COSMIC/pysap) open-source image processing software package has been developed for the 3 years between the Compressed Sensing group at Iniria-CEA Parietal team led by Philippe Ciuciu and the CosmoStat team (CEA/IRFU) led by Jean-Luc Statck. It has been developed for the COmpressed Sensing for Magnetic resonance Imaging and Cosmology (COSMIC) project. This package provides a set of flexible tools that can be applied to a variety of compressed sensing and image reconstruction problems in various research domains. In particular, PySAP offers fast wavelet transforms and a range of integrated optimisation algorithms. It also offers a variety of plugins for specific application domains: on top of Pysap-MRI and PySAP-astro plugins, several complementary modules are now in development for electron tomography and electron microscopy for CEA colleagues. In October 2019, PySAP has been released on PyPi (https://pypi.org/project/python-pySAP/, currently version 0.0.3) and in conda (https://anaconda.org/agrigis/python-pysap).

The Pysap-MRI has been advertised through a specific abstract accepted to the next workshop of ISMRM on Data Sampling & Image Reconstruction in late January 2020. It will be presented during a power pitch session together wih an hands-on demo session using JuPyter notebooks.

- Partner: CEA
- Contact: Philippe Ciuciu

7. New Results

7.1. The visual word form area (VWFA) is part of both language and attention circuitry

While predominant models of visual word form area (VWFA) function argue for its specific role in decoding written language, other accounts propose a more general role of VWFA in complex visual processing. However, a comprehensive examination of structural and functional VWFA circuits and their relationship to behavior has been missing. Here, using high-resolution multimodal imaging data from a large Human Connectome Project cohort (N = 313), we demonstrate robust patterns of VWFA connectivity with both canonical language and attentional networks. Brain-behavior relationships revealed a striking pattern of double dissociation: structural connectivity of VWFA with lateral temporal language network predicted language, but not visuo-spatial attention abilities, while VWFA connectivity with dorsal fronto-parietal attention network predicted visuo-spatial attention, but not language abilities. Our findings support a multiplex model of VWFA function characterized by distinct circuits for integrating language and attention, and point to connectivity-constrained cognition as a key principle of human brain organization.

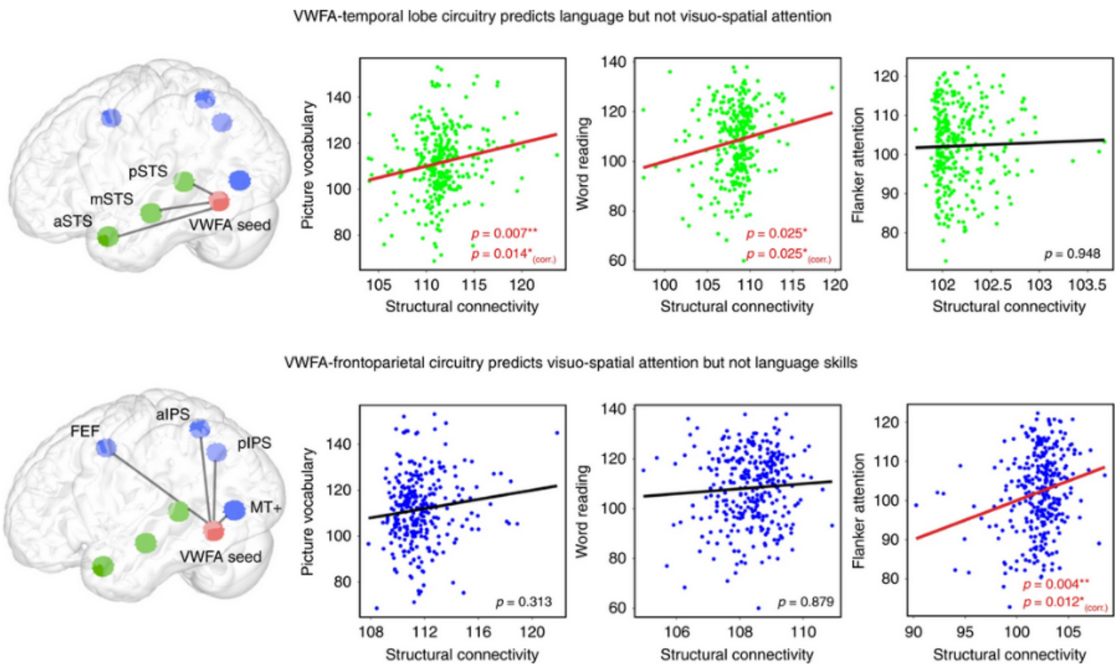


Figure 3. Structural connectivity of VWFA with STS nodes (anterior, middle, and posterior STS) was significantly correlated with individuals' performance on picture vocabulary and word reading tasks, but not on the Flanker visuo-spatial attention task. b Structural connectivity of VWFA with fronto-parietal attention network nodes (FEF, aIPS, pIPS, and MT+) was significantly correlated with individuals' performance on the Flanker attention task but not on the word reading or picture vocabulary tasks. Y-axis is age-adjusted performance scores on cognitive tasks and x-axis presents the predicted performance scores on cognitive tasks from the probability of structural connectivity of VWFA to either the language or the attention network ROIs.

More information can be found in [7].

7.2. SPARKLING: variable-density k-space filling curves for accelerated T2* -weighted MRI

Funding information Purpose: To present a new optimization-driven design of optimal k-space trajectories in the context of compressed sensing: Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING). Theory: The SPARKLING algorithm is a versatile method inspired from stippling techniques that automatically generates optimized sampling patterns compatible with MR hardware constraints on maximum gradient amplitude and slew rate. These non-Cartesian sampling curves are designed to comply with key criteria for optimal sampling: a controlled distribution of samples (e.g., variable density) and a locally uniform k-space coverage. Methods: Ex vivo and in vivo prospective $T2^{\star}$-weighted acquisitions were performed on a 7 Tesla scanner using the SPARKLING tra-jectories for various setups and target densities. Our method was compared to radial and variable-density spiral trajectories for high resolution imaging. Results: Combining sampling efficiency with compressed sensing, the proposed sampling patterns allowed up to 20-fold reductions in MR scan time (compared to fully-sampled Cartesian acquisitions) for two-dimensional $T2^{\star}$-weighted imaging without deterioration of image quality, as demonstrated by our experimental results at 7 Tesla on in vivo human brains for a high in-plane resolution of 390 um. In comparison to existing non-Cartesian sampling strategies, the proposed technique also yielded superior image quality. Conclusion: The proposed optimization-driven design of k-space trajectories is a versatile framework that is able to enhance MR sampling performance in the context of compressed sensing.

More information can be found in [14].

7.3. Benchmarking functional connectome-based predictive models for resting-state fMRI

Functional connectomes reveal biomarkers of individual psychological or clinical traits. However, there is great variability in the analytic pipelines typically used to derive them from rest-fMRI cohorts. Here, we consider a specific type of studies, using predictive models on the edge weights of functional connectomes, for which we highlight the best modeling choices. We systematically study the prediction performances of models in 6 different cohorts and a total of 2 000 individuals, encompassing neuro-degenerative (Alzheimer's, Post-traumatic stress disorder), neuro-psychiatric (Schizophrenia, Autism), drug impact (Cannabis use) clinical settings and psychological trait (fluid intelligence). The typical prediction procedure from rest-fMRI consists of three main steps: defining brain regions, representing the interactions, and supervised learning. For each step we benchmark typical choices: 8 different ways of defining regions –either pre-defined or generated from the rest-fMRI data– 3 measures to build functional connectomes from the extracted time-series, and 10 classification models to compare functional interactions across subjects. Our benchmarks summarize more than 240 different pipelines and outline modeling choices that show consistent prediction performances in spite of variations in the populations and sites. We find that regions defined from functional data work best; that it is beneficial to capture between-region interactions with tangent-based parametrization of covariances, a midway between correlations and partial correlation; and that simple linear predictors such as a logistic regression give the best predictions. Our work is a step forward to establishing reproducible imaging-based biomarkers for clinical settings.

More information can be found in [8].

7.4. Population shrinkage of covariance (PoSCE) for better individual brain functional-connectivity estimation

Estimating covariances from functional Magnetic Resonance Imaging at rest (r-fMRI) can quantify interactions between brain regions. Also known as brain functional connectivity, it reflects inter-subject variations in behavior and cognition, and characterizes neuropathologies. Yet, with noisy and short time-series, as in r-fMRI, covariance estimation is challenging and calls for penalization, as with shrinkage approaches. We

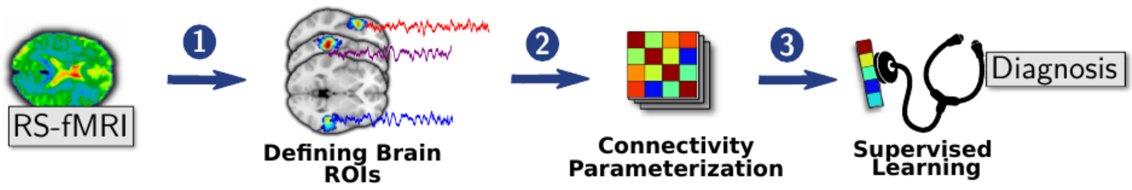


Figure 4. Functional connectome-prediction pipeline with three main steps: 1) definition of brain regions (ROIs) from rest-fMRI images or us- ing already defined reference atlases, 2) quantifying functional interactions from time series signals extracted from these ROIs and 3) comparisons of functional interactions across subjects using supervised learning.

introduce population shrinkage of covariance estimator (PoSCE) : a covariance estimator that integrates prior knowledge of covariance distribution over a large population, leading to a non-isotropic shrinkage. The shrinkage is tailored to the Riemannian geometry of symmetric positive definite matrices. It is coupled with a probabilistic modeling of the individual and population covariance distributions. Experiments on two large r-fMRI datasets (HCP n=815, Cam-CAN n=626) show that PoSCE has a better bias-variance trade-off than existing covariance estimates: this estimator relates better functional-connectivity measures to cognition while capturing well intra-subject functional connectivity.

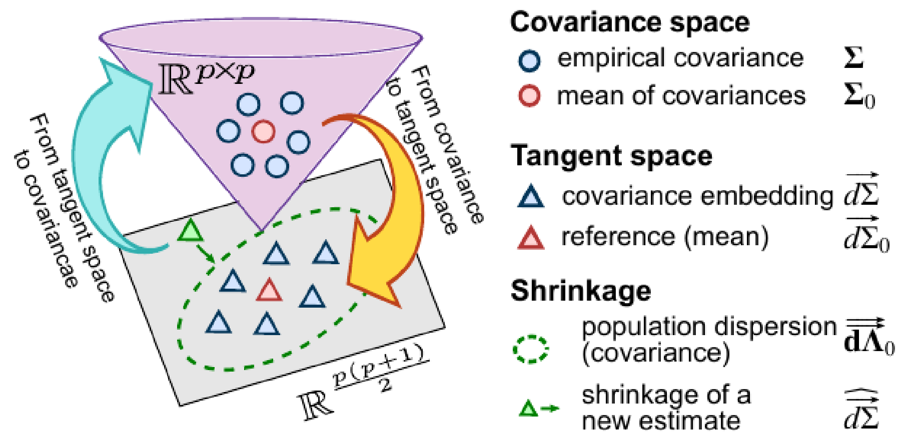


Figure 5. Tangent embedding and population prior modeling. Σ_0 is the mean covariance from a train set of covariances. It is the reference point in the tangent space. The population prior is defined as a Gaussian multivariate dis- tribution centered on $d\Sigma_0$. Λ_0 is the covariance dispersion over the population. The arrows depict the mapping between the non-Euclidean covariance space and the tangent space.

More information can be found in [20].

7.5. Feature Grouping as a Stochastic Regularizer for High-Dimensional Structured Data

In many applications where collecting data is expensive , for example neuroscience or medical imaging, the sample size is typically small compared to the feature dimension. It is challenging in this setting to train expressive, non-linear models without overfitting. These datasets call for intelligent regularization that exploits known structure, such as correlations between the features arising from the measurement device. However, existing structured regularizers need specially crafted solvers, which are difficult to apply to complex models. We propose a new regularizer specifically designed to leverage structure in the data in a way that can be applied efficiently to complex models. Our approach relies on feature grouping, using a fast clustering algorithm inside a stochas-tic gradient descent loop: given a family of feature groupings that capture feature covariations, we randomly select these groups at each iteration. We show that this approach amounts to enforcing a denoising regularizer on the solution. The method is easy to implement in many model archi-tectures, such as fully connected neural networks, and has a linear computational cost. We apply this regularizer to a real-world fMRI dataset and the Olivetti Faces datasets. Experiments on both datasets demonstrate that the proposed approach produces models that generalize better than those trained with conventional regularizers, and also improves convergence speed.

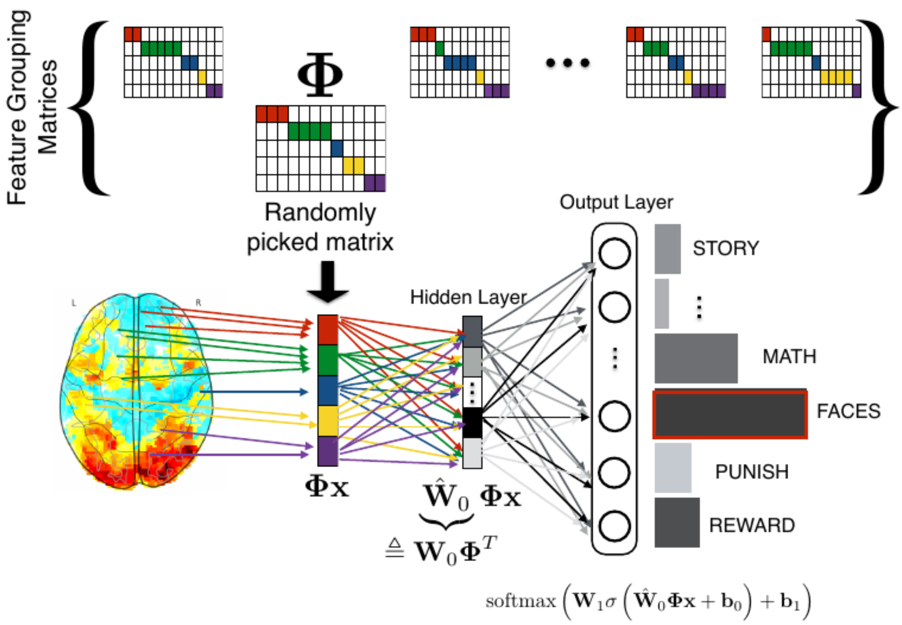


Figure 6. Illustration of the proposed approach: Forward propa- gation of a neural network with a single hidden layer using feature grouping during training. The parameters of the neural network to be estimated are $\mathbf{W}_0$, $\mathbf{b}_0$, $\mathbf{W}_1$, $\mathbf{b}1$. A bank of feature grouping matrices are pre-generated where each matrix is calculated from a sub-sample of the training test. At each SGD iteration, a feature grouping matrix is sampled from the bank of pre-generated matri- ces. The gradient is then computed with respect to $\widehat{\mathbf{W}}_0$ to update $\mathbf{W}_0$ in backpropagation.

More information can be found in [25].

7.6. Manifold-regression to predict from MEG/EEG brain signals without source modeling

Magnetoencephalography and electroencephalography (M/EEG) can reveal neuronal dynamics non-invasively in real-time and are therefore appreciated methods in medicine and neuroscience. Recent advances in modeling brain-behavior relationships have highlighted the effectiveness of Riemannian geometry for summarizing the spatially correlated time-series from M/EEG in terms of their covariance. However , after artefact-suppression, M/EEG data is often rank deficient which limits the application of Riemannian concepts. In this article, we focus on the task of regression with rank-reduced covariance matrices. We study two Riemannian approaches that vectorize the M/EEG covariance between-sensors through projection into a tangent space. The Wasserstein distance readily applies to rank-reduced data but lacks affine-invariance. This can be overcome by finding a common sub-space in which the covariance matrices are full rank, enabling the affine-invariant geometric distance. We investigated the implications of these two approaches in synthetic generative models, which allowed us to control estimation bias of a linear model for prediction. We show that Wasserstein and geometric distances allow perfect out-of-sample prediction on the generative models. We then evaluated the methods on real data with regard to their effectiveness in predicting age from M/EEG covariance matrices. The findings suggest that the data-driven Riemannian methods outperform different sensor-space estimators and that they get close to the performance of biophysics-driven source-localization model that requires MRI acquisitions and tedious data processing. Our study suggests that the proposed Riemannian methods can serve as fundamental building-blocks for automated large-scale analysis of M/EEG.

Figure 7. Age prediction on Cam-CAN MEG dataset for different methods, ordered by out-of-sample MAE. The y-axis depicts the projection method, with identity denoting the absence of projection. Colors indicate the subsequent embedding. The biophysics-driven MNE method (blue) performs best. The Riemannian methods (orange) follow closely and their performance depends little on the projection method. The non-Riemannian methods log-diag (green) perform worse, although the supervised projection clearly helps.

More information can be found in [47].

7.7. Stochastic algorithms with descent guarantees for ICA

Independent component analysis (ICA) is a widespread data exploration technique, where observed signals are modeled as linear mixtures of independent components. From a machine learning point of view, it amounts to a matrix factorization problem with a statistical independence criterion. Infomax is one of the most used ICA algorithms. It is based on a loss function which is a non-convex log-likelihood. We develop a new majorization-minimization framework adapted to this loss function. We derive an online algorithm for the streaming setting, and an incremental algorithm for the finite sum setting, with the following benefits. First, unlike most algorithms found in the literature, the proposed methods do not rely on any critical hyper-parameter like a step size, nor do they require a line-search technique. Second, the algorithm for the finite sum setting, although stochas-tic, guarantees a decrease of the loss function at each iteration. Experiments demonstrate progress on the state-of-the-art for large scale datasets, without the necessity for any manual parameter tuning.

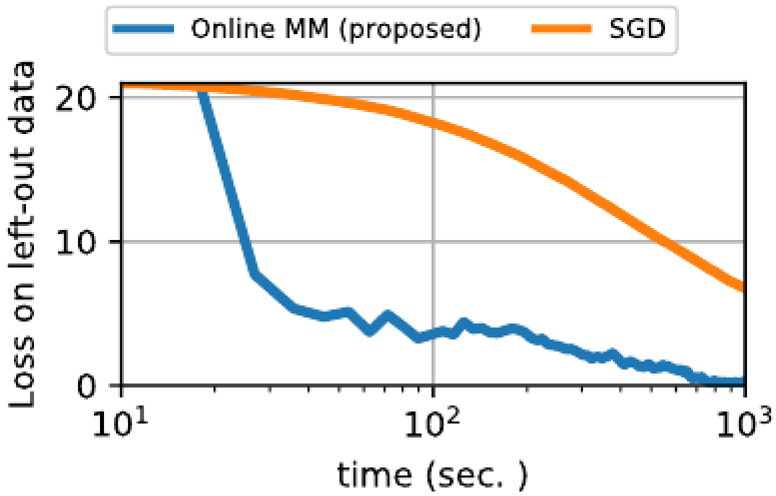


Figure 8. Online algorithms applied on a 32 GB real dataset with p = 100 and n = $4{\times}10^7$ *. Time is in logarithmic scale. Values of the loss on left out data greater than its initial value are truncated.*

More information can be found in [39].

7.8. Comparing distributions: $l1$ geometry improves kernel two-sample testing

Are two sets of observations drawn from the same distribution? This problem is a two-sample test. Kernel methods lead to many appealing properties. Indeed state-of-the-art approaches use the L2 distance between kernel-based distribution representatives to derive their test statistics. Here, we show that Lp distances (with $p \geq 1$) between these distribution representatives give metrics on the space of distributions that are well-behaved to detect differences between distributions as they metrize the weak convergence. Moreover, for analytic kernels, we show that the L1 geometry gives improved testing power for scalable computational procedures. Specifically, we derive a finite dimensional approximation of the metric given as the ℓ_1 norm of a vector which captures differences of expectations of analytic functions evaluated at spatial locations or frequencies (i.e, features). The features can be chosen to maximize the differences of the distributions and give interpretable indications of how they differs. Using an ℓ_1 norm gives better detection because differences between representatives are dense as we use analytic kernels (non-zero almost everywhere). The tests are consistent, while much faster than state-of-the-art quadratic-time kernel-based tests. Experiments on artificial and real-world problems demonstrate improved power/time tradeoff than the state of the art, based on ℓ_2

norms, and in some cases, better outright power than even the most expensive quadratic-time tests. More information can be found in [37].

7.9. Wasserstein regularization for sparse multi-task regression

We focus in this work on high-dimensional regression problems where each regressor can be associated to a location in a physical space, or more generally a generic geometric space. Such problems often employ sparse priors, which promote models using a small subset of regressors. To increase statistical power, the so-called multi-task techniques were proposed, which consist in the simultaneous estimation of several related models. Combined with sparsity assumptions, it lead to models enforcing the active regressors to be shared across models, thanks to, for instance L1 / Lq norms. We argue in this paper that these techniques fail to leverage the spatial information associated to regressors. Indeed, while sparse priors enforce that only a small subset of variables is used, the assumption that these regressors overlap across all tasks is overly simplistic given the spatial variability observed in real data. In this paper, we propose a convex reg-ularizer for multi-task regression that encodes a more flexible geometry. Our regularizer is based on unbalanced optimal transport (OT) theory, and can take into account a prior geometric knowledge on the regressor variables, without necessarily requiring overlapping supports. We derive an efficient algorithm based on a regularized formulation of OT, which iterates through applications of Sinkhorn's algorithm along with coordinate descent iterations. The performance of our model is demonstrated on regular grids with both synthetic and real datasets as well as complex triangulated geometries of the cortex with an application in neuroimaging.

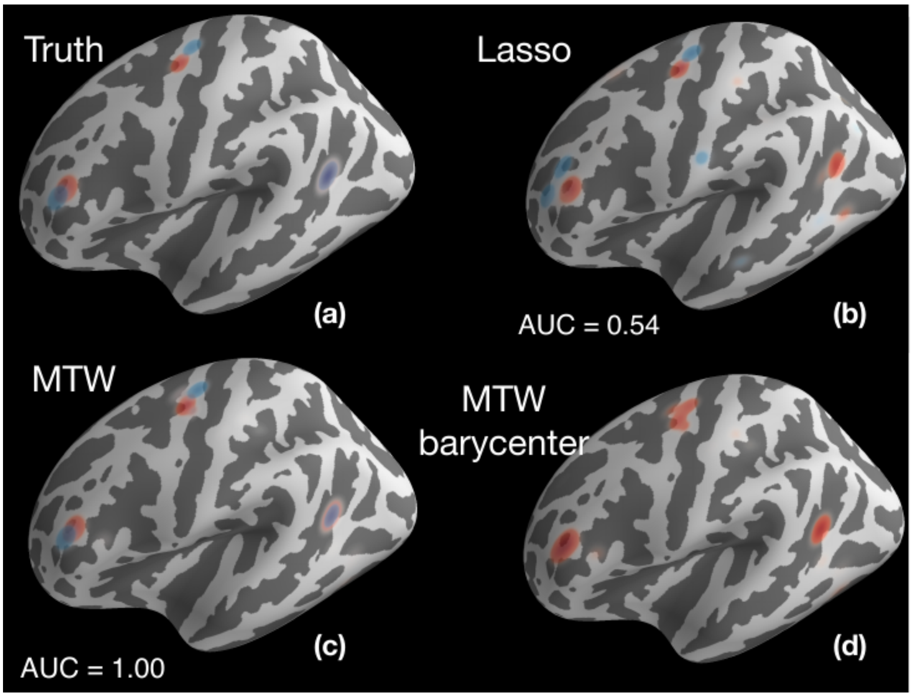


Figure 9. Each color corresponds to one of the two subjects (except for (d)). (a): True sources: one common feature (right side of the displayed hemisphere) and two non-overlapping sources. (b, c): Sources estimated by (b) Lasso and (c) MTW with the highest AUC score. (d) Shows the barycenter texttheta¯ associated with MTW model. In this figure, the displayed activations were smoothed for the sake of visibility.

More information can be found in [34].

7.10. Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data

Continuous improvement in medical imaging techniques allows the acquisition of higher-resolution images. When these are used in a predictive setting, a greater number of explanatory variables are potentially related to the dependent variable (the response). Meanwhile, the number of acquisitions per experiment remains limited. In such high dimension/small sample size setting, it is desirable to find the explanatory variables that are truly related to the response while controlling the rate of false discoveries. To achieve this goal, novel multivariate inference procedures, such as knockoff inference, have been proposed recently. However, they require the feature covariance to be well-defined, which is impossible in high-dimensional settings. In this paper, we propose a new algorithm, called Ensemble of Clustered Knockoffs, that allows to select explanatory variables while controlling the false discovery rate (FDR), up to a prescribed spatial tolerance. The core idea is that knockoff-based inference can be applied on groups (clusters) of voxels, which drastically reduces the problem's dimension; an ensembling step then removes the dependence on a fixed clustering and stabilizes the results. We benchmark this algorithm and other FDR-controlling methods on brain imaging datasets and observe empirical gains in sensitivity, while the false discovery rate is controlled at the nominal level.

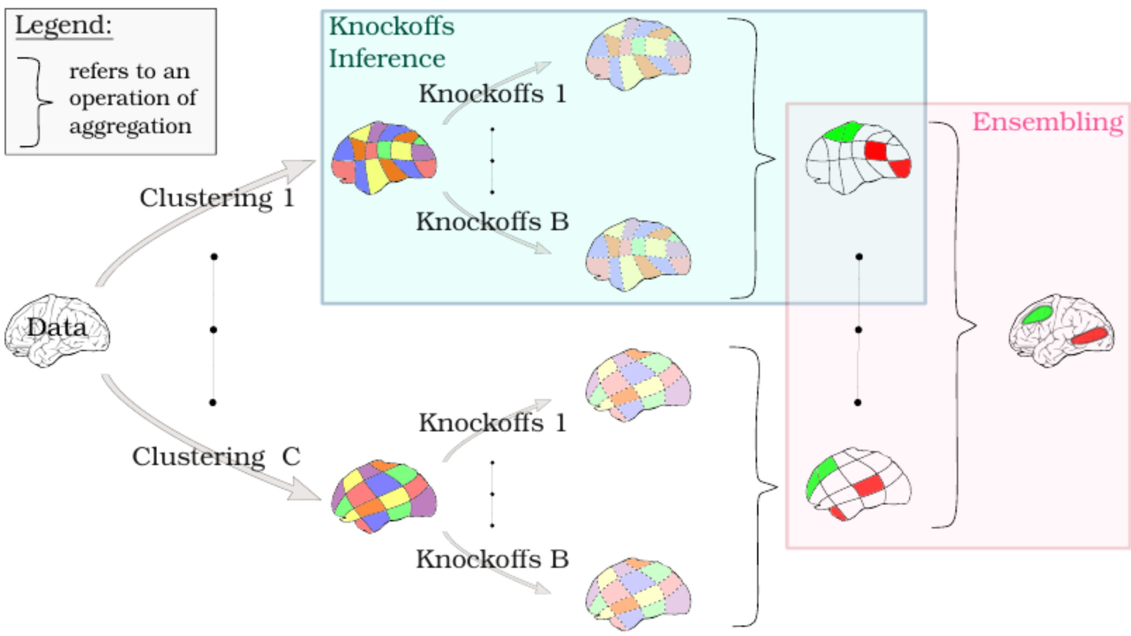


Figure 10. Representation of the ECKO algorithm. To create a stable inference result, we introduce ensembling steps both within each cluster level and at the voxel- level, across clusterings.

More information can be found in [46].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

In 2019, a CIFRE PhD thesis was launched with Siemens-Healthineers France. This contract supports the PhD thesis of Guillaume Daval-Frérot.

8.2. Scikit-learn Consortium

Scikit-learn is a machine-learning library in Python. It is the engine that powers many applications of artifical intelligence and data science.

Scikit-learn is used on a regular basis by more than half a million people in the world, with applications ranging from medical imaging to product recommendation.

Scikit-learn is an open-source software, under BSD license that facilitates commercial usage. It is developed by a world-wide community, gathering many different expertise on statistics, algorithms and software production.

The quality of scikit-learn, its algorithms, its interfaces, its documentation, are universally acclaimed. Its development follows a strict process to ensure this quality.

The goal of the foundation is to enable maintaining scikit-learn's high standards addressing new challenges.

The foundation employs central contributors to the project, to support scikit-learn's community and to develop new ambitious features. The priorities of the foundation are set jointly by the community and its sponsors.

More information can be found here http://scikit-learn.fondation-inria.fr/home.

The consortium is supported by 8 companies and has an annual budget of about half a million euros.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Inserm-Inria project

This project is funded by the joint Inserm and Inria program 'médecine numérique' and is conducted in collaborations with our clinical partners from the Lariboisière hospital, Inserm uni U942 BioCANVAS (Biomarkers in Cardio-Neuro-VAScular diseases). It supports the PhD thesis of David Sabbagh.

Participants:

- Denis Engemann [coordinator, co-advisor]
- Alexandre Gramfort [thesis director, co-advisor]
- Etienne Gayat [clinical collaborator, co-advisor]
- Fabrice Vallée [clinical collaborator]
- David Sabbagh [PhD Student]

Post-operative delirium (POD) is a potential complication of anesthesia during surgery. It is often associated with adverse outcomes and is aggravated by aging. In elderly patients, post-operative complications have been estimated to incur tens of million US dollars of costs each year in the United States by prolonging hospitalization and potentially affecting health prognosis. Recent studies suggest that POD can already be prevented by improving electrophysiological monitoring of anesthesia depth and individual dosage of anesthetic agents. Doing so probably minimizes the time patients spend in a coma-like state that manifests itself in isoelectric burst suppression, an electroencephalogram (EEG) pattern characterized by alternation between quiescence and high-amplitude bursts, and causally linked to POD. However, such an enterprise, currently, depends on the trained clinical electrophysiologist and guidance by commercially provided EEG indices of states of consciousness. One such metric is the bispectral index (BIS), which, like other related metrics, does not explicitly take into account baseline changes related to normative aging and may therefore be biased when used naively.

While electrophysiological signatures of aging (e.g. drop in Alpha and Gamma band power), states of consciousness (e.g. drop in Theta band long-range connectivity) and drug response (e.g. anteriorization of alpha band power in propofol anesthesia) have been separately investigated in the past years, their common denominators are not known. It is therefore difficult to detect individual risk, choose the optimal dosage, and automate anesthesia monitoring readily for any patient in any hospital.

The goal of this research project is to build statistical models that enable prediction of burst suppression and subsequent POD by exploiting diverse EEG-signatures of states of consciousness in the context of aging. We approach this challenge by recasting it as a problem of learning brain-age from the point of view of electrophysiology of consciousness.

9.1.2. CoSmic project

Participants: Philippe Ciuciu [Correspondant], Nicolas Chartier, Loubna El Gueddari, Zaccharie Ramzi, Chaithya Giliyar Radhkrishna.

This project is funded by CEA DRF-Impulsion.

the DRF-impulsion CEA program which has been transformed into a CEA PTC program for 2 years (2018-2020), in collaboration with Pierre Kestener, La Maison de la Simulation (CEA/CNRS).

Compressed Sensing is a recent theory in maths that allows the perfect recovery of signals or images from compressive acquisition scenarios. This approach has been popularized in MRI over the last decade as well as in astrophysics (noticeably in radio-astronomy). So far, both of these fields have developed skills in CS separately. The aim of the COSMIC project is to foster collaborations between CEA experts in MRI (Parietal team within NeuroSpin) and in astrophysics (CosmoStat lab within the Astrophysics Department). These interactions will allow us to share different expertise in order to improve image quality, either in MRI or in radio-astronomy (thanks to the interferometry principle). In this field, given the data delivered by radio-telescopes, the goal consists in extracting high temporal resolution information in order to study fast transient events.

9.1.3. Metacog

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Jérome Dockès.

This project is funded by Digiteo.

This is a Digicosme project (2016-2019) and a collaboration with Fabian Suchanek (Telecom Paritech).

Understanding how cognition emerges from the billions of neurons that constitute the human brain is a major open problem in science that could bridge natural science –biology– to humanities –psychology. Psychology studies performed on humans with functional Magnetic Resonance Imaging (fMRI) can be used to probe the full repertoire of high-level cognitive functions. While analyzing the resulting image data for a given experiment is a relatively well-mastered process, the challenges in comparing data across multiple datasets poses serious limitation to the field. Indeed, such comparisons require to pool together brain images acquired under different settings and assess the effect of different *experimental conditions* that correspond to psychological effects studied by neuroscientists.

Such meta-analyses are now becoming possible thanks to the development of public data resources –OpenfMRI http://openfmri.org and NeuroVault http://neurovault.org. As many others, researchers of the Parietal team understand these data sources well and contribute to them. However, in such open-ended context, the description of experiments in terms of cognitive concepts is very difficult: there is no universal definition of cognitive terms that could be employed consistently by neuroscientists. Hence meta-analytic studies loose power and specificity. On the other hand, http://brainspell.org provide a set of curated annotation, albeit on much less data, that can serve as a seed or a ground truth to define a consensual ontology of cognitive concepts. Relating these terms to brain activity poses another challenge, of statistical nature, as brain patterns form high-dimensional data in perspective with the scarcity and the noise of the data.

The purpose of this project is to learn a semantic structure in cognitive terms from their occurrence in brain activation. This structure will simplify massive multi-label statistical-learning problems that arise in brain mapping by providing compact representations of cognitive concepts while capturing the imprecision on the definition these concepts.

9.1.4. HidimStat

Participants: Bertrand Thirion [Correspondant], Jerome-Alexis Chevalier, Joseph Salmon.

This project is funded by Digiteo.

This is a Digicosme project (2017-2020) and a collaboration with Joseph Salmon (Telecom Paritech).

The HiDimStat project aims at handling uncertainty in the challenging context of high dimensional regression problem. Though sparse models have been popularized in the last twenty years in contexts where many features can explain a phenomenon, it remains a burning issue to attribute confidence to the predictive models that they produce. Such a question is hard both from the statistical modeling point of view, and from a computation perspective. Indeed, in practical settings, the amount of features at stake (possibly up to several millions in high resolution brain imaging) limit the application of current methods and require new algorithms to achieve computational efficiency. We plan to leverage recent developments in sparse convex solvers as well as more efficient reformulations of testing and confidence interval estimates to provide several communities with practical software handling uncertainty quantification. Specific validation experiments will be performed in the field of brain imaging.

9.1.5. Template estimation for arbitrary alignments: application to brain imaging.

Participants: Bertrand Thirion [Correspondant], Thomas Bazeille.

This project is funded by Digiteo.

In the recent years, the nature of scientific inference has shifted quite substantially from model-based to predictive approaches, thanks to the generalization of powerful machine learning techniques. While this has certainly improved scientific standards, this has also obscured the objects and concepts on which inference is drawn. For instance, it is now possible –based on some initial data– to predict individual brain activity topographies, yet the very notion of a standard brain template has become increasingly elusive. Given the importance of establishing models for the progress of knowledge, we revisit the problem of model inference on data with high variance. Specifically, in a context where almost arbitrary transformation can successfully warp observations to each other with high accuracy, what is the common definition of a population model underlying all these observations? What is the working definition of a template ? We plan to leverage recent developments on optimal transport and multivariate analysis to build working definition of templates; we will use them in a brain imaging context to build a novel generation of brain templates.

9.1.6. CDS2

Participants: Alexandre Gramfort [Correspondant], Gaël Varoquaux, Maria Telenczuk, Jiaping Liu.

CDS2 is an "Strategic research initiative" of the Paris Saclay University Idex https://www.datascience-paris-saclay.fr/. Although it groups together many partners of the Paris Saclay ecosystem, Parietal has been deeply involved in the project. It currently funds 2 engineers: Maria Telenczuk and Jiaping (Lucy) Liu.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. Neuroref: Mathematical Models of Anatomy / Neuroanatomy / Diffusion MRI

Participants: Demian Wassermann [Correspondant], Antonia Machlouzarides Shalit, Valentin Iovene.

While mild traumatic brain injury (mTBI) has become the focus of many neuroimaging studies, the understanding of mTBI, particularly in patients who evince no radiological evidence of injury and yet experience clinical and cognitive symptoms, has remained a complex challenge. Sophisticated imaging tools are needed to delineate the kind of subtle brain injury that is extant in these patients, as existing tools are often ill-suited for the diagnosis of mTBI. For example, conventional magnetic resonance imaging (MRI) studies have focused on seeking a spatially consistent pattern of abnormal signal using statistical analyses that compare average differences between groups, i.e., separating mTBI from healthy controls. While these methods are successful in many diseases, they are not as useful in mTBI, where brain injuries are spatially heterogeneous.

The goal of this proposal is to develop a robust framework to perform subject-specific neuroimaging analyses of Diffusion MRI (dMRI), as this modality has shown excellent sensitivity to brain injuries and can locate subtle brain abnormalities that are not detected using routine clinical neuroradiological readings. New algorithms will be developed to create Individualized Brain Abnormality (IBA) maps that will have a number of clinical and research applications. In this proposal, this technology will be used to analyze a previously acquired dataset from the INTRuST Clinical Consortium, a multi-center effort to study subjects with Post-Traumatic Stress Disorder (PTSD) and mTBI. Neuroimaging abnormality measures will be linked to clinical and neuropsychological assessments. This technique will allow us to tease apart neuroimaging differences between PTSD and mTBI and to establish baseline relationships between neuroimaging markers, and clinical and cognitive measures.

9.2.1.2. DirtyData: Data integration and cleaning for statistical analysis

Participants: Gaël Varoquaux [Correspondant], Patricio Cerda Reyes, Pierre Glaser.

Machine learning has inspired new markets and applications by extracting new insights from complex and noisy data. However, to perform such analyses, the most costly step is often to prepare the data. It entails correcting errors and inconsistencies as well as transforming the data into a single matrix-shaped table that comprises all interesting descriptors for all observations to study. Indeed, the data often results from merging multiple sources of informations with different conventions. Different data tables may come without names on the columns, with missing data, or with input errors such as typos. As a result, the data cannot be automatically shaped into a matrix for statistical analysis.

This proposal aims to drastically reduce the cost of data preparation by integrating it directly into the statistical analysis. Our key insight is that machine learning itself deals well with noise and errors. Hence, we aim to develop the methodology to do statistical analysis directly on the original dirty data. For this, the operations currently done to clean data before the analysis must be adapted to a statistical framework that captures errors and inconsistencies. Our research agenda is inspired from the data-integration state of the art in database research combined with statistical modeling and regularization from machine learning.

Data integrating and cleaning is traditionally performed in databases by finding fuzzy matches or overlaps and applying transformation rules and joins. To incorporate it in the statistical analysis, an thus propagate uncertainties, we want to revisit those logical and set operations with statistical-learning tools. A challenge is to turn the entities present in the data into representations well-suited for statistical learning that are robust to potential errors but do not wash out uncertainty.

Prior art developed in databases is mostly based on first-order logic and sets. Our project strives to capture errors in the input of the entries. Hence we formulate operations in terms of similarities. We address typing entries, deduplication -finding different forms of the same entity- building joins across dirty tables, and correcting errors and missing data.

Our goal is that these steps should be generic enough to digest directly dirty data without user-defined rules. Indeed, they never try to build a fully clean view of the data, which is something very hard, but rather include in the statistical analysis errors and ambiguities in the data.

The methods developed will be empirically evaluated on a variety of dataset, including the French public-data repository, http://www.data.gouv.fr. The consortium comprises a company specialized in data integration, Data Publica, that guides business strategies by cross-analyzing public data with market-specific data.

9.2.1.3. FastBig Project

Participants: Bertrand Thirion [Correspondant], Jerome-Alexis Chevalier, Tuan Binh Nguyen.

In many scientific applications, increasingly-large datasets are being acquired to describe more accurately biological or physical phenomena. While the dimensionality of the resulting measures has increased, the number of samples available is often limited, due to physical or financial limits. This results in impressive amounts of complex data observed in small batches of samples.

A question that arises is then : what features in the data are really informative about some outcome of interest ? This amounts to inferring the relationships between these variables and the outcome, conditionally to all other variables. Providing statistical guarantees on these associations is needed in many fields of data science, where competing models require rigorous statistical assessment. Yet reaching such guarantees is very hard.

FAST-BIG aims at developing theoretical results and practical estimation procedures that render statistical inference feasible in such hard cases. We will develop the corresponding software and assess novel inference schemes on two applications : genomics and brain imaging.

9.2.1.4. MultiFracs project

Participant: Philippe Ciuciu [Correspondant].

The scale-free concept formalizes the intuition that, in many systems, the analysis of temporal dynamics cannot be grounded on specific and characteristic time scales. The scale-free paradigm has permitted the relevant analysis of numerous applications, very different in nature, ranging from natural phenomena (hydrodynamic turbulence, geophysics, body rhythms, brain activity,...) to human activities (Internet traffic, population, finance, art,...).

Yet, most successes of scale-free analysis were obtained in contexts where data are univariate, homogeneous along time (a single stationary time series), and well-characterized by simple-shape local singularities. For such situations, scale-free dynamics translate into global or local power laws, which significantly eases practical analyses. Numerous recent real-world applications (macroscopic spontaneous brain dynamics, the central application in this project, being one paradigm example), however, naturally entail large multivariate data (many signals), whose properties vary along time (non-stationarity) and across components (non-homogeneity), with potentially complex temporal dynamics, thus intricate local singular behaviors.

These three issues call into question the intuitive and founding identification of scale-free to power laws, and thus make uneasy multivariate scale-free and multifractal analyses, precluding the use of univariate methodologies. This explains why the concept of scale-free dynamics is barely used and with limited successes in such settings and highlights the overriding need for a systematic methodological study of multivariate scale-free and multifractal dynamics. The Core Theme of MULTIFRACS consists in laying the theoretical foundations of a practical robust statistical signal processing framework for multivariate non homogeneous scale-free and multifractal analyses, suited to varied types of rich singularities, as well as in performing accurate analyses of scale-free dynamics in spontaneous and task-related macroscopic brain activity, to assess their natures, functional roles and relevance, and their relations to behavioral performance in a timing estimation task using multimodal functional imaging techniques.

This overarching objective is organized into 4 Challenges:

1. Multivariate scale-free and multifractal analysis,
2. Second generation of local singularity indices,
3. Scale-free dynamics, non-stationarity and non-homogeneity,
4. Multivariate scale-free temporal dynamics analysis in macroscopic brain activity.

9.2.1.5. DARLING: Distributed adaptation and learning over graph signals

Participant: Philippe Ciuciu [Correspondant].

The project will be starting in 2020 with a post-doc to be hired probably in 2021.

The DARLING project will aim to propose new adaptive learning methods, distributed and collaborative on large dynamic graphs in order to extract structured information of the data flows generated and/or transiting at the nodes of these graphs. In order to obtain performance guarantees, these methods will be systematically accompanied by an in-depth study of random matrix theory. This powerful tool , never exploited so far in this context although perfectly suited for inference on random graphs, will thereby provide even avenues for improvement. Finally, in addition to their evaluation on public data sets, the methods will be compared with each other using two advanced imaging techniques in which two of the partners are involved: radio astronomy with the giant SKA instrument (Obs. Côte d'Azur) and magnetoencephalographic brain imaging (Inria Parietal at NeuroSpin, CEA Saclay). These involve the processing of time series on graphs while operating at extreme observation scales.

9.2.1.6. meegBIDS.fr: Standardization, sharing and analysis of MEEG data simplified by BIDS

Participant: Alexandre Gramfort [Correspondant].

The project accepted by ANR in 2019 will be starting in 2020 with an engineer to be hired in 2020. This project is in collaboration with the MEG groups at CEA NeuroSpin and the Brain and Spine Institute (ICM) in Paris.

The neuroimaging community recently started an international effort to standardize the sharing of data recorded with magnetoencephalography (MEG) and with electroencephalography (EEG). This format, known as the Brain Imaging Data Structure (BIDS), now needs a wider adoption, notably in the French neuroimaging community, along with the development of dedicated software tools that operate seamlessly on BIDS formatted datasets. The meegBIDS.fr project has three aims: 1) accelerate the research cycles by allowing analysis software tools to work with BIDS formated data, 2) simplify data sharing with high quality standards thanks to automated validation tools, 3) train French neuroscientists to leverage existing public BIDS MEG/EEG datasets and to share their own data with little efforts.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. VirtualBrainCloud

Title:

Programm: H2020 FET Open

Duration: 01/01/2019 - 31/12/2022

Coordinator: Petra Ritter

Inria contact: Bertrand Thirion

Summary:

The central goal of this project is the development of a cloud-based platform for biomedical research and clinical decision-making that helps to improve early patient-specific diagnosis and treatment of NDD and has substantial potential for significant positive socioeconomic impact.

The platform integrates several aims that revolve around early diagnosis, prognosis, and personalized treatment of neurodegenerative diseases (NDD) like Alzheimer's disease (AD) and Parkinson's disease (PD). It is becoming increasingly clear that meeting this objective requires a multifactorial approach that takes into account individual genetic, metabolic and environmental aspects, and that integrates them with the understanding of the biophysical processes underlying NDD.

More information can be found here https://virtualbraincloud-2020.eu/tvb-cloud-main.html.

9.3.1.2. Neurolang

Title: Accelerating Neuroscience Research by Unifying Knowledge Representation and Analysis Through a Domain Specific Language

Programm: ERC Starting researcher

Duration: 01/03/2018 - 28/02/2023

Coordinator: Demian Wassermann

Inria contact: Demian Wassermann

Summary:

Neuroscience is at an inflection point. The 150-year old cortical specialization paradigm, in which cortical brain areas have a distinct set of functions, is experiencing an unprecedented momentum with over 1000 articles being published every year. However, this paradigm is reaching its limits. Recent studies show that current approaches to atlas brain areas, like relative location, cellular population type, or connectivity, are not enough on their own to characterize a cortical area and its function unequivocally. This hinders the reproducibility and advancement of neuroscience.

Neuroscience is thus in dire need of a universal standard to specify neuroanatomy and function: a novel formal language allowing neuroscientists to simultaneously specify tissue characteristics, relative location, known function and connectional topology for the unequivocal identification of a given brain region.

The vision of NeuroLang is that a unified formal language for neuroanatomy will boost our understanding of the brain. By defining brain regions, networks, and cognitive tasks through a set of formal criteria, researchers will be able to synthesize and integrate data within and across diverse studies. NeuroLang will accelerate the development of neuroscience by providing a way to evaluate anatomical specificity, test current theories, and develop new hypotheses.

NeuroLang will lead to a new generation of computational tools for neuroscience research. In doing so, we will be shedding a novel light onto neurological research and possibly disease treatment and palliative care. Our project complements current developments in large multimodal studies across different databases. This project will bring the power of Domain Specific Languages to neuroscience research, driving the field towards a new paradigm articulating classical neuroanatomy with current statistical and machine learning-based approaches.

9.3.1.3. SLAB (698)

Title: Signal processing and Learning Applied to Brain data

Programm: ERC Starting researcher

Duration: 01/04/2017 - 31/08/2021

Coordinator: Alexandre Gramfort

Partner: LTCI , Telecom ParisTech (France)

Inria contact: Alexandre Gramfort

Summary:

Understanding how the brain works in healthy and pathological conditions is considered as one of the challenges for the 21st century. After the first electroencephalography (EEG) measurements in 1929, the 90's was the birth of modern functional brain imaging with the first functional MRI and full head magnetoencephalography (MEG) system. In the last twenty years, imaging has revolutionized clinical and cognitive neuroscience.

After pioneering works in physics and engineering, the field of neuroscience has to face two major challenges. The size of the datasets keeps growing. The answers to neuroscience questions are limited by the complexity of the signals observed: non-stationarity, high noise levels, heterogeneity of sensors, lack of accurate models. SLAB will provide the next generation of models and algorithms for mining electrophysiology signals which offer unique ways to image the brain at a millisecond time scale.

SLAB will develop dedicated machine learning and signal processing methods and favor the emergence of new challenges for these fields. SLAB focuses on five objectives: 1) source localization with M/EEG for brain imaging at high temporal resolution 2) representation learning to boost statistical power and reduce acquisition costs 3) fusion of heterogeneous sensors 4) modeling of non-stationary spectral interactions to identify functional coupling between neural ensembles 5) development of fast algorithms easy to use by non-experts.

SLAB aims to strengthen mathematical and computational foundations of brain data analysis. The methods developed will have applications across fields (computational biology, astronomy, econometrics). Yet, the primary impact of SLAB will be on neuroscience. The tools and high quality open software produced in SLAB will facilitate the analysis of electrophysiology data, offering new perspectives to understand how the brain works at a mesoscale, and for clinical applications (epilepsy, autism, tremor, sleep disorders).

9.3.1.4. HBP SGA2

Title: Interactive Computing E-Infrastructure for the Human Brain Project

Programm: FET Flagship

Duration: 01/04/2018 - 31/03/2020

Coordinator: Katrin Amunts

Partners: see https://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners/

Inria contact: Bertrand Thirion

Summary:

The HBP Flagship was launched by the European Commission's Future and Emerging Technologies (FET) scheme in October 2013, and is scheduled to run for ten years. The Flagships, represent a new partnering model for visionary, long-term European cooperative research in the European Research Area, demonstrating the potential for common research efforts. The HBP has the following main objectives:

- Create and operate a European scientific Research Infrastructure for brain research, cognitive neuroscience, and other brain-inspired sciences
- Gather, organise and disseminate data describing the brain and its diseases
- Simulate the brain
- Build multi-scale scaffold theory and models for the brain
- Develop brain-inspired computing, data analytics and robotics
- Ensure that the HBP's work is undertaken responsibly and that it benefits society.

More information on the HBP's Flagship Objectives is available in the Framework Partnership Agreement.

The timeline of the Project is split into multiple phases, each of which will be covered by a separate funding agreement. The current phase is Specific Grant Agreement Two (SGA2), which spans the two-year period from April 2018–April 2020. The HBP is funded via several sources. Total funding is planned to be in the region of EUR 1 billion; around one half of which will be provided by the European Union, and the other by Member States and private funding sources. The European Union contributed EUR 54 million to the Project in the Ramp-Up Phase (October 2013 to March 2016), EUR 89 million for the second phase (SGA1), and EUR 88 million for the current phase (SGA2). The FET Flagships Staff Working Document provides further information on how Flagships are funded.

9.4. International Initiatives

9.4.1. Inria International Labs

Inria@SiliconValley

Associate Team involved in the International Lab:

9.4.1.1. Meta&Co

Title: Meta-Analysis of Neuro-Cognitive Associations

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Psychology department. - Russel Poldrack

Start year: 2018

See also: http://team.inria.fr/parietal

Cognitive science and psychiatry describe mental operations: cognition, emotion, perception and their dysfunction. Cognitive neuroimaging bridge these mental concepts to their implementation in the brain, neural firing and wiring, by relying on functional brain imaging. Yet aggregating results from experiments probing brain activity into a consistent description faces the roadblock that cognitive concepts and brain pathologies are ill-defined. Separation between them is often blurry. In addition, these concepts and subdivisions may not correspond to actual brain structures or systems. To tackle this challenge, we propose to adapt data-mining techniques used to learn relationships in computational linguistics. Natural language processing uses distributional semantics to build semantic relationships and ontologies. New models are needed to learn relationships from heterogeneous signals: functional magnetic resonance images (fMRI), on the one hand, combined with related psychology and neuroimaging annotations or publications, on the other hand. Such a joint effort will rely on large publicly-available fMRI databases shared by Podrack Lab, as well as literature mining.

Inria@SiliconValley

Associate Team involved in the International Lab:

9.4.1.2. LargeSmallBrainNets

Title: Characterizing Large and Small-scale Brain Networks in Typical Populations Using Novel Computational Methods for dMRI and fMRI-based Connectivity and Microstructure

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Stanford Cognitive and Systems Neuroscience Laboratory - Vinod Menon

Start year: 2019

See also: http://pages.saclay.inria.fr/demian.wassermann/largesmallbrainnets/

In the past two decades, brain imaging of neurotypical individuals and clinical populations has primarily focused on localization of function and structures in the brain, revealing activation in specific brain regions during performance of cognitive tasks through modalities such as functional MRI. In parallel, technologies to identify white matter structures have been developed using diffusion MRI. Lately, interest has shifted towards developing a deeper understanding of the brain's macroscopic and microscopic architectures and their influence on cognitive and affective information processing. Using for this resting state fMRI and diffusion MRI to build the functional and structural networks of the human brain.

The human brain is a complex patchwork of interconnected regions, and graph-theoretical approaches have become increasingly useful for understanding how functionally connected systems engender, and constrain, cognitive functions. The functional nodes of the human brain, i.e. cortical regions, and their structural inter-connectivity, collectively the brain's macrostructure or "connectome", are, however, poorly understood. Quantifying in vivo how these nodes' microstructure, specifically cellular composition or cytoarchitecture, influences the cognitive tasks in which these are involved is fundamental problem in understanding the connectome. Furthermore, the coupling between within and across-subject contributions to the connectome and cognitive differences hampers the identification and understanding of the link between brain structure and function, and human cognition.

Critically, there is a dearth of computational methods for reliably identifying functional nodes of the brain, their micro and macrostructure in vivo, and separating the population and subject-specific effects. Devising and validating methods for investigating the human connectome has therefore taken added significance.

The first major goal of this project is to develop and validate appropriate sophisticated computational and mathematical tools relate the brain's macrostructure with its function. Specifically, we will focus on being able to separate population and subject-specific contributions within these models using state-of-the-art human brain imaging techniques and open-source data from the Human Connectome Project (HCP) and the Adolescent Brain Cognitive Development study (ABCD). To this end, we will first develop and validate novel computational tools for (1) formulating and fitting large scale random effect models on graphs derived from functional and structural connectivity and (2) implement techniques enabling us to impose different regularization schemes based on sparsity and multicollinearity of the model parameters.

The second major goal of this project is characterizing the cytoarchitecture of the nodes, i.e. cortical regions, at the microscopic level and their relationship with the brain's hemodynamical function and cognition. For this, we will (1) identify cortical areas with specific cytoarchitecture in the human cortex and use them to develop diffusion MRI-based models, (2) validate these models with numerical simulations of the dMRI signal and animal models, and (3) establish the relationship between cytoarchitecture and hemodynamical function measured from fMRI and cognition. For this we will leverage multi-shell high-angular diffusion MRI from public databases such as HCP and ABCD.

Finally, we will use to use our newly developed computational tools to characterize normal structural and functional brain networks in neurotypical adults. Due to the complementarity of the cognitive science and imaging techniques expertise the synergy between the two laboratories of this associate team will allow us to reveal in unprecedented detail the structural and functional connectivity of the human brain and its relation to cognition.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Aapo Havärinen has been with parietal since April 2019 for a one-year visit, funded by the Dataia convergence institute.
- Luigi Gresele (MPI Tübingen) has been visiting the team in November-December 2019.
- Cedric Xu (UPenn) visited the team for three months in June-August 2019.
- James Cole (UCL) visited the team in December 2019.
- Yu Zhang (BIC, Montreal) visited the team during one week in September 2019
- Bemsibom Toh (HWU, Edinburgh) visited the team during two months in September-November 2019

9.5.2. Visits to International Teams

9.5.2.1. Sabbatical programme

Gael Varoquaux is spending one year in Montreal (September 2019-September 2020), hosted at MILA and Montreal Neuroimaging Institute at McGill Unviversity.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

Bertrand Thirion co-organized the "Imagerie médicale et intelligence artificielle " colloque at Collège de France on April, 23rd, 2019.

Thomas Moreau co-organized the Palaisien seminar: a monthly scientific seminar in statistics and statistical learning bringing together researchers in this field working on the Saclay Plateau. The seminar started in November and brings together about fifty researchers (http://palaisien.herokuapp.com).

10.1.1.2. Member of the Organizing Committees

Demian Wassermann: MICCAI 2019, ISMRM 2019

Gaël Varoquaux: PyParis 2019

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- Philippe Ciuciu: Chairman of scientific oral sessions in ISBI 2019 (Venice, Italy), program commitee of EUSIPCO 2019.
- Gaël Varoquaux: program committee of NIPS, ICML, ICLR, senior program committee IJCAI.
- Alexandre Gramfort: program committee of NIPS, ICML, ICLR.
- Demian Wassermann: ISMRM 2019
- Thomas Moreau: JDSE 2019

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Philippe Ciuciu: Senior Area Editor of IEEE OJSP
- Alexandre Gramfort: Editor, NeuroImage, Journal of Machine Learning Research (JMLR), Frontiers in Brain Imaging Methods
- Bertrand Thirion: Editor, Frontiers in Brain Imaging Methods
- Gael Varoquaux: Editor, elife

10.1.3.2. Reviewer - Reviewing Activities

- Philippe Ciuciu is reviewer for Nature Communication, Biological Pysichiatry, Plos Computational Biology, Neuroimage, Scientific Reports, Journal of Neuroscience, IEEE TSP/TIP/TMI/TCI/SPL, Frontiers in Neuroscience, MRM, SIAM Imaging Science
- Gaël Varoquaux: Nature Methods, JMLR, PLOS Bio, NeuroImage, NeuroImage Clinical, IEEE TMI, Annals of Applied Statistics, Biological Psychiatry, MedIA, Science
- Alexandre Gramfort: JMLR, PLOS Computational Biology, NeuroImage, IEEE TBME, IEEE TMI, IEEE TSP, MedIA, NIPS, ICML, ICLR, ICASSP, Scientific Reports, Frontiers in Brain Imaging Methods, Journal of Neuroscience Methods
- Bertrand Thirion: Nature communications, Neuroimage, Medical Image Analysis, IEEE TMI, PNAS, PLOS Comp Bio, Brain Structure and Function, NIPS, ICML, IPMI, AISTATS
- Denis Engemann: Brain, PLOS Biology, PLOS Computational Biology, Psychological Medicine, Scientific Reports, Neuroimage, Neuroimage Clinical, Human Brain Mapping, Journal of Machine Learning Research, Brain Topography, Brain Connectivity, Journal of Alzheimer's Disease, Neuroscience of Consciousness, PLOS ONE, Frontiers in Neuroscience, Journal of Computational Neuroscience, Psychiatry and Clinical Neurosciences, Sensors
- Demian Wassermann: NeuroImage, MRM, JMRI, Brain Structure and Function, Cortex, MedIA.
- Thomas Moreau: ICML, NeurIPS, EUSIPCO, JDSE, AISTAT, JMLR, TSP, LSP, JMIV
- Olivier Grisel: JMLR

10.1.4. Invited Talks

- Philippe Ciuciu, Ecole Centrale Nantes, Nov. 2019
- Philippe Ciuciu, I3S/CNRS (Sophia-Antipolis), Oct. 2019

- Philippe Ciuciu, Wavelets & Sparsity XVIII, San Diego, Aug. 2019
- Philippe Ciuciu, Collège de France, Apr. 2019
- Denis Engemann, Max Planck Institute AES, Frankfurt, Germany, Jan. 2019
- Denis Engemann, University of Birmingham, Birmingham, UK, Apr. 2019
- Denis Engemann, University of Glasgow, Glasgow, UK, Jun. 2019
- Denis Engemann, Max Planck Institute CBS, Leipzig, Germany, Nov. 2019
- Alexandre Gramfort at IHP, Paris, France, April. 2019
- Alexandre Gramfort at Univ Genoa, Italy, Sept. 2019
- Alexandre Gramfort at Univ Jyvaskyla, Finland, May. 2019
- Alexandre Gramfort at Institut Pasteur, Paris, France, June 2019
- Alexandre Gramfort at National Institute of Health, Washington, USA, Oct. 2019
- Bertrand Thirion at Donders Institute, Netherlands Feb. 2019
- Bertrand Thirion at MRC CBU, Cambridge, UK, Nov. 2019
- Bertrand Thirion at INT, Marseille, Nov. 2019
- Bertrand Thirion at GFAIH, Paris, Oct. 2019
- Bertrand Thirion at Université Paris-Sud, faculté de médecine, Nov. 2019
- Bertrand Thirion at Alan turing Institute, London, June 2019
- Gael Varoquaux at MNI (Montréal), Nov. 2019
- Gael Varoquaux at SparkAISummit, Amsterdam, Oct. 2019
- Gael Varoquaux at PyData Amsterdam, Oct. 2019
- Gael Varoquaux at Mila (Montréal), Sept. 2019
- Gael Varoquaux at Etalab (Paris), June 2019
- Gael Varoquaux at MGH (Harvard), June 2019
- Gael Varoquaux at Yale, June 2019
- Demian Wassermann at NYU, February 2019
- Demian Wassermann at ERC, June 2019
- Demian Wassermann at BWH (Harvard), June 2019
- Demian Wassermann at Stanford, October 2019
- Thomas Moreau **ISMRM, Educational Session, Montreal, Canada, 2019:** "Best Practices & Pitfalls in Applying Machine Learning to Magnetic Resonance Imaging".
- Thomas Moreau **Télécom ParisTech, SMILE seminar, Paris, France, 2019:** "Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals".
- Thomas Moreau **ENS Paris-Saclay, Centre de Mathématiques et de leurs Applications, MLMDA seminar, Cachan, France, 2019:** "Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals".
- Thomas Moreau **Telecom ParisTech, Laboratoire Traitement et Communication de l'Information, seminar, Paris, France, 2019:** "Using the Dictionary Structure for efficient Convolutional Dictionary Learning".
- Thomas Moreau **Centre pour l'Énergie Atomique Saclay, Cosmostat, annual seminar day, Saclay, France, 2019:** "Learning Recurring Patterns in Large Signals with Convolutional Dictionary Learning".

10.1.5. Leadership within the Scientific Community

- Demian Wassermann: Organising committee of the BrainHack community — Paris Chapter

- Demian Wassermann: ISMRM and MICCAI action organization towards reducing gender and minority biases.

10.1.6. Scientific Expertise

- Philippe Ciuciu has been reviewer for NWO, The Netherlands (Veni/Vidi/Vici call)
- Philippe Ciuciu has been hired as H2020 expert evaluator for the H2020-WIDESPREAD-2020-5 Twinning RCA programme since Dec 2019.
- Denis Engemann has been reviewer for the Swiss National Science Foundation in 2019
- Alexandre Gramfort has been member of an NSF panel in the USA in January 2019.
- Demian Wassermann FET-OPEN ERCA Action, and DataIA
- Gaël Varoquaux was expert for DataIA

10.1.7. Research Administration

- In Nov 2018, Philippe Ciuciu Elected Vice-Chair of the SAT Biomedical Image & Signal Analytics (EURASIP technical committee) for 2019-2020.
- In Dec 2019, Philippe Ciuciu has been renewed as the IEEE Signal Processing Society Representative for the 2020 IEEE ISBI conference.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Alexandre Gramfort, Optimization for Data Science, 20h, Msc 2 Data Science Master Ecole Polytechnique, France

Master: Alexandre Gramfort, DataCamp, 20h, Msc 2 Data Science Master Ecole Polytechnique, France

Master: Alexandre Gramfort, Source Imaging with EEG and MEG, 7h, Msc 2 in Biomedical Imaging at Télécom Paristech

Doctoral School: Alexandre Gramfort, Practical machine learning, 6h, The mathematics of imaging: the CIRM pre-school, Marseille, France

Master: Bertrand Thirion, Functional neuroimaging and BCI, 12h, Master MVA, ENS Paris-Saclay, France

Master: Bertrand Thirion, Kshitih Chawla, Thomas Bazeille: tutorial on Nilearn, 8h, INT, Marseille.

Master: Philippe Ciuciu, fMRI course: From acquisition to data Analysis, 6h, Msc 2 in Biomedical Imaging, Université Paris-Sud

Bachelor: Demian Wassermann, CSE201 class, 15h, C++ programming, Ecole Polytechnique

Master: Demian Wassermann, 7h Biomedical Engineering, Msc 2 Biomedical Engineering

Extension: Demian Wassermann, Data Science, 20h, Ecole Polytechnique

Master: Gaël Varoquaux, Machine learning in Python, 3h, ENSAE

Master: Gaël Varoquaux, Functional brain connectivity, 7h, Msc 2 in Biomedical Imaging at Télécom Paristech

Doctoral school: Gaël Varoquaux, representation learning with limited sample, 4h30, Deep Learning Summer School, Warsaw

Doctoral school: Gaël Varoquaux, machine learning for neuroimaging, 3h, Unique days, Montréal

Doctoral school: Gaël Varoquaux, machine learning with scikit-learn, 4h, IBPRIA, Madrid

Master: Thomas Moreau, **Formation continue Telecom ParisTech (7h)** TP pour Machine learning and Python.

Master: Thomas Moreau, **Master Spécialisé courses, Telecom ParisTech (20h)** Assistant TP pour le cours de machine learning avancé.

Master: Thomas Moreau, **Exectutive Master Statistique et Big Data, Université Paris-Dauphine (9h)** Cours Python pour le machine learning.

Master: Thomas Moreau, **Master Data Science, Université Paris-Saclay (25h)** Cours machine learning associé à un challenge.

Master: Olivier Grisel, **Master Data Science, Université Paris-Saclay (40h)** Cours de Deep Learning.

10.2.2. Supervision

Demian Wassermann has defended his habilitation thesis in May 2019.

The following PhD defenses have taken place

Guillermo Gallardo, "Inferring and comparing structural parcellations of the human brain using diffusion MRI", defended Jan 2019, co-directed by Demian Wassermann

Patricio Cerda, "encoding high-cardinality string categories", defended Nov 2019, directed by Gaël Varoquaux

Jérome Dockes, "large-scale predictive neuroimaging meta-analyses" defended Nov 2019, directed by Gaël Varoquaux & Fabian Suchanek (Telecom)

Pierre Ablin, defended Nov 2019, directed by Alexandre Gramfort

Mathurin Massias defended Dec 2019, directed by Alexandre Gramfort

Loubna El Guedari, defended Dec 2019, directed by Philippe Ciuciu

Ongoing PhD theses

Kamalakar Daddi, "predictive population analysis for mental health", defense planned Oct 2020, directed by Gaël Varoquaux, Denis Engemmann & Bertrand Thirion

Alexis Cvetkov Iliev, "embeddings of database entries", defense planned Oct 2022, directed by Gaël Varoquaux & Alexandre Allauzen (ESPCI)

Hicham Janati, 01/10/2017, coadvised by Alexandre Gramfort and Marco Cuturi

Quentin Bertrand, 01/10/2018, coadvised by Alexandre Gramfort and Joseph Salmon

Hubert Banville, 01/10/2018, coadvised by Alexandre Gramfort and Denis Engemann

David Sabbagh, 01/10/2018, coadvised by Alexandre Gramfort and Denis Engemann

Jérôme-Alexis Chevalier, 01/10/2017, coadvised by Bertrand Thirion and Joseph Salmon

Hugo Richard, 01/10/2018, coadvised by Bertrand Thirion and Olivier Grisel

Thomas Bazeille, 01/10/2018, advised by Bertrand Thirion

Tuan Binh Nguyen, 01/10/2018, coadvised by Bertrand thirion and Sylvain Arlot

Valentin Iovene, 01/10/2018, advised by Demian Wassermann

Antonia Machlouzarides Shalit, 01/10/2018, coadvised by Demian Wassermann and Bertrand Thirion

Hamza Cherkaoui, 01/10/2017, co-advised by Philippe Ciuciu and Thomas Moreau

Zaccharie Ramzi, 01/02/2019, co-advised by Philippe Ciuciu and Jean-Luc Starck

Anaïïs Artiges, 01/10/2019, co-advised by Philippe Ciuciu and Cyril Poupon

Guillaume Daval-Frérot, 01/11/2019, advised by Philippe Ciuciu

Maeliss Jallais, 01/10/2018, advised by Demian Wassermann

Lucas Martin, 01/10/2019, advised by Bertrand Thirion and Julie Josse.

Chengran Fan, 01/10/2019, co-advised by Demian Wassermann and Jing-Rebecca Li

Gaston Zanitti, 01/11/2019, advised by Demian Wassermann

10.2.3. Juries

Bertrand Thirion has been involved in the following committees:

- PhD - Florian Tilquin / University of Strasbourg
- PhD - Han Bossier / University of Ghent
- Habilitation - Laurent Oudre / Université Paris XIII

Philippe Ciuciu has been involved in the following committees:

- PhD - Marica Pesce (reviewer) / Heriot-Watta University, Edinburgh
- PhD - Christian El Hajj (reviewer) / Ecole Centrale Nantes.

Denis Engemann has been involved in the following committees:

- PhD - Laura Sophie Imperatori, IMT School for Advanced Studies Lucca, Lucca, Italy
- MD - Cyril Touchard, Universite´ Paris Diderot (Paris 7)

Gael Varoquaux has been involved in the following committees:

- PhD - Beyrem KHALFAOUI / Mines ParisTech

Alexandre Gramfort has been involved in the following committees:

- PhD - Benjamin Donnot / LRI, Université Paris-Sud
- PhD - Belhal Karimi / CMAP, Polytechnique
- PhD -Pedro Rodriges (reviewer) / GIPSA Lab, Grenoble
- PhD - Kostiantyn Maksymenko (reviewer) / Inria, Sophia
- MD - Cyril Touchard, Universite´ Paris Diderot (Paris 7)

Demian Wassermann has been involved in the following committees:

- PhD-midterm - Katrin Rojkova / Université Paris-Sorbonne
- Habilitation - Daniel Margulies / Université Paris-Sorbonne

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Philippe Ciuciu has been member of the Inria Saclay scientific commission since 2016
- Denis Engemann has been managing the Twitter accounts of Parietal and the MNE software since 2017
- Alexandre Gramfort is in charge of the Paris-Saclay Center for Data Science
- Alexandre is a Member of the technical development committee of Inria Saclay since Dec 2018
- Bertrand Thirion, head of the Dataia Convergence Institute since January 1st, 2019
- Gaël Varoquaux was Member of the technical development committee of Inria Saclay until Dec 2018
- Gaël Varoquaux is member of the doctoral monitoring committee of Inria Saclay
- Gaël Varoquaux is member of the steering committee of the Paris-Saclay Center for Data Science
- Thomas Moreau: Co-organisation of the weekly Parietal team seminar bringing together the 45 members of the team for short presentations and a longer scientific presentation.
- Gaël Varoquaux is in charge of the scikit-learn consortium

10.3.2. Internal action

- Guillaume Lemaître, Marie Telenczuk, et Lucie Liu have taught a 1-day training on data-science with Python in Nov 2019.

10.3.3. Creation of media or tools for science outreach

- Guillaume Lemaître, Gael Varoquaux, Olivier Grisel started developing new teaching material for MOOC of Scikit-learn to be hosted in 2020 on the FUN platform in collaboration with Loic Esteve and the team members of the Inria Learning Lab).

11. Bibliography

Major publications by the team in recent years

[1] A. MENSCH, J. MAIRAL, B. THIRION, G. VAROQUAUX.*Stochastic Subsampling for Factorizing Huge Matrices*, in "IEEE Transactions on Signal Processing", January 2018, vol. 66, n^o^ 1, p. 113-128 [*DOI :* 10.1109/TSP.2017.2752697], https://hal.archives-ouvertes.fr/hal-01431618

Publications of the year

Doctoral Dissertations and Habilitation Theses

[2] M. MASSIAS.*Sparse high dimensional regression in the presence of colored heteroscedastic noise: application to M/EEG source imaging*, Telecom Paristech, December 2019, https://tel.archives-ouvertes.fr/tel-02401628

Articles in International Peer-Reviewed Journal

[3] D. BZDOK, D. FLORIS, A. MARQUAND.*Analyzing Brain Networks in Population Neuroscience: A Case for the Bayesian Philosophy*, in "Philosophical Transactions of the Royal Society of London. B (1887–1895)", January 2020, https://hal.archives-ouvertes.fr/hal-02447507

[4] D. BZDOK, J. P. A. IOANNIDIS.*Exploration, inference and prediction in neuroscience and biomedicine*, in "Trends in Neurosciences", March 2019, https://arxiv.org/abs/1903.10310 [*DOI :* 10.1016/J.TINS.2019.02.001], https://hal.archives-ouvertes.fr/hal-02044120

[5] D. BZDOK, T. E. NICHOLS, S. SMITH.*Towards Algorithmic Analytics for Large-scale Datasets*, in "Nature Machine Intelligence", July 2019 [*DOI :* 10.1038/S42256-019-0069-5], https://hal.archives-ouvertes.fr/hal-02178410

[6] S. CHAMBON, V. THOREY, P. J. ARNAL, E. J. M. MIGNOT, A. GRAMFORT.*DOSED: A deep learning approach to detect multiple sleep micro-events in EEG signal*, in "Journal of Neuroscience Methods", June 2019, vol. 321, p. 64-78, https://arxiv.org/abs/1812.04079 [*DOI :* 10.1016/J.JNEUMETH.2019.03.017], https://hal.inria.fr/hal-02121090

[7] L. CHEN, D. WASSERMANN, D. ABRAMS, J. KOCHALKA, G. GALLARDO-DIEZ, V. MENON.*The visual word form area (VWFA) is part of both language and attention circuitry*, in "Nature Communications", December 2019, vol. 10, n^o^ 1, Lang Chen, Demian Wassermann, and Daniel Abrams contributed equally [*DOI :* 10.1038/S41467-019-13634-Z], https://hal.inria.fr/hal-02401938

[8] K. DADI, M. RAHIM, A. ABRAHAM, D. CHYZHYK, M. MILHAM, B. THIRION, G. VAROQUAUX.*Benchmarking functional connectome-based predictive models for resting-state fMRI*, in "NeuroImage", May 2019, n^o^ 192, p. 115-134 [*DOI :* 10.1016/J.NEUROIMAGE.2019.02.062], https://hal.inria.fr/hal-01824205

[9] R. H. FICK, D. WASSERMANN, R. DERICHE.*The Dmipy Toolbox: Diffusion MRI Multi-Compartment Modeling and Microstructure Recovery Made Easy*, in "Frontiers in Neuroinformatics", October 2019, vol. 13 [*DOI :* 10.3389/FNINF.2019.00064], https://hal.archives-ouvertes.fr/hal-02400877

[10] P. FILIPIAK, R. H. FICK, A. PETIET, M. SANTIN, A.-C. PHILIPPE, S. LEHÉRICY, P. CIUCIU, R. DERICHE, D. WASSERMANN.*Reducing the number of samples in spatiotemporal dMRI acquisition design*, in "Magnetic Resonance in Medicine", 2019 [*DOI :* 10.1002/MRM.27601], https://hal.archives-ouvertes.fr/hal-01928734

[11] T. HE, R. KONG, A. J. HOLMES, M. Q. NGUYEN, M. SABUNCU, S. B. EICKHOFF, D. BZDOK, J. FENG, B. T. YEO.*Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics*, in "NeuroImage", October 2019, 116276 [*DOI :* 10.1016/J.NEUROIMAGE.2019.116276], https://hal.archives-ouvertes.fr/hal-02314718

[12] M. JORDING, D. ENGEMANN, H. ECKERT, G. BENTE, K. VOGELEY.*Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues*, in "Frontiers in Human Neuroscience", December 2019, vol. 13 [*DOI :* 10.3389/FNHUM.2019.00442], https://hal.archives-ouvertes.fr/hal-02416981

[13] T. M. KARRER, D. S. BASSETT, B. DERNTL, O. GRUBER, A. ALEMAN, R. JARDRI, A. R. LAIRD, P. FOX, S. B. EICKHOFF, O. GRISEL, G. VAROQUAUX, B. THIRION, D. BZDOK.*Brain-based ranking of cognitive domains to predict schizophrenia*, in "Human Brain Mapping", July 2019 [*DOI :* 10.1002/HBM.24716], https://hal.archives-ouvertes.fr/hal-02180311

[14] C. LAZARUS, P. WEISS, N. CHAUFFERT, F. MAUCONDUIT, L. EL GUEDDARI, C. DESTRIEUX, I. ZEMMOURA, A. VIGNAUD, P. CIUCIU.*SPARKLING: variable-density k-space filling curves for accelerated T 2 * -weighted MRI*, in "Magnetic Resonance in Medicine", June 2019, vol. 81, n^{o} 6, p. 3643-3661 [*DOI :* 10.1002/MRM.27678], https://hal.inria.fr/hal-02361265

[15] J. LEBENBERG, J.-F. MANGIN, B. THIRION, C. POUPON, L. HERTZ-PANNIER, F. LEROY, P. ADIBPOUR, G. DEHAENE-LAMBERTZ, J. DUBOIS.*Mapping the asynchrony of cortical maturation in the infant brain: a MRI multi-parametric clustering approach*, in "NeuroImage", January 2019, vol. 185, p. 641-653 [*DOI :* 10.1016/J.NEUROIMAGE.2018.07.022], https://hal.archives-ouvertes.fr/hal-01966812

[16] J. LI, T. BOLT, D. BZDOK, J. S. NOMI, B. T. YEO, R. NATHAN SPRENG, L. Q. UDDIN.*Topography and Behavioral Relevance of the Global Signal in the Human Brain*, in "Scientific Reports", October 2019, vol. 9, n^{o} 1 [*DOI :* 10.1038/S41598-019-50750-8], https://hal.archives-ouvertes.fr/hal-02305188

[17] G. NGUYEN, N. AUNAI, D. FONTAINE, E. LE PENNEC, J. VAN DEN BOSSCHE, A. JEANDET, B. BAKKALI, L. VIGNOLI, B. REGALDO-SAINT BLANCARD.*Automatic Detection of Interplanetary Coronal Mass Ejections from In Situ Data: A Deep Learning Approach*, in "The Astrophysical journal letters", April 2019, vol. 874, n^{o} 2, 145, https://arxiv.org/abs/1903.10780 [*DOI :* 10.3847/1538-4357/AB0D24], https://hal.sorbonne-universite.fr/hal-02103805

[18] V.-D. NGUYEN, M. DE LEONI, T. DANCHEVA, J. JANSSON, J. HOFFMAN, D. WASSERMANN, J.-R. LI.*Portable simulation framework for diffusion MRI*, in "Journal of Magnetic Resonance", December 2019, vol. 309, 106611 [*DOI :* 10.1016/J.JMR.2019.106611], https://hal.archives-ouvertes.fr/hal-02431598

[19] L. J. O'DONNELL, A. DADUCCI, D. WASSERMANN, C. LENGLET.*Advances in computational and statistical diffusion MRI*, in "NMR in Biomedicine", 2019, vol. 32, n^{o} 4, e3805 [*DOI :* 10.1002/NBM.3805], https://hal.inria.fr/hal-02432249

[20] M. RAHIM, B. THIRION, G. VAROQUAUX.*Population shrinkage of covariance (PoSCE) for better individual brain functional-connectivity estimation*, in "Medical Image Analysis", March 2019 [*DOI :* 10.1016/J.MEDIA.2019.03.001], https://hal.inria.fr/hal-02068389

[21] S. SAREMI, A. HYVÄRINEN.*Neural Empirical Bayes*, in "Journal of Machine Learning Research", 2019, vol. 20, p. 1 - 23, https://hal.inria.fr/hal-02419496

[22] R. TOMI-TRICOT, V. GRAS, B. THIRION, F. MAUCONDUIT, N. BOULANT, H. CHERKAOUI, P. ZERBIB, A. VIGNAUD, A. LUCIANI, A. AMADON.*SmartPulse, a Machine Learning Approach for Calibration-Free Dynamic RF Shimming: Preliminary Study in a Clinical Environment*, in "Magnetic Resonance in Medicine", June 2019, vol. 82, p. 2016–2031 [*DOI :* 10.1002/MRM.27870], https://hal-cea.archives-ouvertes.fr/cea-02141266

[23] G. VAROQUAUX, R. POLDRACK.*Predictive models avoid excessive reductionism in cognitive neuroimaging*, in "Current Opinion in Neurobiology", April 2019, vol. 55, forthcoming [*DOI :* 10.1016/J.CONB.2018.11.002], https://hal.archives-ouvertes.fr/hal-01856412

International Conferences with Proceedings

[24] P. ABLIN, D. FAGOT, H. WENDT, A. GRAMFORT, C. FÉVOTTE.*A Quasi-Newton Algorithm on the Orthogonal Manifold for NMF with Transform Learning*, in "IEEE-ICASSP 2019 - International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, https://hal.archives-ouvertes.fr/hal-02346829

[25] S. AYDÖRE, B. THIRION, G. VAROQUAUX.*Feature Grouping as a Stochastic Regularizer for High-Dimensional Structured Data*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, https://hal.inria.fr/hal-02318458

[26] T. BIN MASOOD, J. BUDIN, M. FALK, G. FAVELIER, C. GARTH, C. GUEUNET, P. GUILLOU, L. HOFMANN, P. HRISTOV, A. KAMAKSHIDASAN, C. KAPPE, P. KLACANSKY, P. LAURIN, J. A. LEVINE, J. LUKASCZYK, D. SAKURAI, M. SOLER, P. STENETEG, J. TIERNY, W. USHER, J. VIDAL, M. WOZNIAK.*An Overview of the Topology ToolKit*, in "TopoInVis 2019 - Topological Methods in Data Analysis and Visualization", Nykoping, Sweden, June 2019, https://hal.archives-ouvertes.fr/hal-02159838

[27] H. CHERKAOUI, T. MOREAU, A. HALIMI, P. CIUCIU.*Sparsity-based blind deconvolution of neural activation signal in fMRI*, in "IEEE-ICASSP 2019 - International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, May 2019, https://hal.inria.fr/hal-02085810

[28] O. D. DOMINGUES, P. CIUCIU, D. LA ROCCA, P. ABRY, H. WENDT.*Multifractal analysis for cumulant-based epileptic seizure detection in eeg time series*, in "ISBI 2019 - IEEE International Symposium on Biomedical Imaging", Venise, Italy, April 2019, https://hal.inria.fr/hal-02108099

[29] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*Calibrationless oscar-based image reconstruction in compressed sensing parallel MRI*, in "ISBI 2019 - IEEE International Symposium on Biomedical Imaging", Venise, Italy, April 2019, https://hal.inria.fr/hal-02101262

[30] P. FILIPIAK, R. FICK, A. PETIET, M. SANTIN, A.-C. PHILIPPE, S. LEHÉRICY, R. DERICHE, D. WASSERMANN.*Coarse-Grained Spatiotemporal Acquisition Design for Diffusion MRI*, in "ISBI 2019 - IEEE International Symposium on Biomedical Imaging", Venice, Italy, April 2019, https://hal.inria.fr/hal-01973588

[31] L. E. GUEDDARI, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET, P. CIUCIU.*Online MR image reconstruction for compressed sensing acquisition in T2* imaging*, in "SPIE Conference - Wavelets and Sparsity XVIII", San Diego, United States, August 2019, https://hal.inria.fr/hal-02265538

[32] A. HALIMI, P. CIUCIU, A. MCCARTHY, S. MCLAUGHLIN, G. S. BULLER.*Fast adaptive scene sampling for single-photon 3D lidar images*, in "IEEE CAMSAP 2019 - International Workshop on Computational Advances in Multi-Sensor Adaptive Processing", Le Gosier (Guadeloupe), France, December 2019, https://hal.inria.fr/hal-02298998

[33] H. JANATI, T. BAZEILLE, B. THIRION, M. CUTURI, A. GRAMFORT.*Group level MEG/EEG source imaging via optimal transport: minimum Wasserstein estimates*, in "IPMI 2019 - 26th international conference on Information Processing in Medical Imaging", Hong Kong, Hong Kong SAR China, SPRINGER (editor), June 2019, vol. Lecture Notes in Computer Science, https://arxiv.org/abs/1902.04812 , https://hal.inria.fr/hal-02013889

[34] H. JANATI, M. CUTURI, A. GRAMFORT.*Wasserstein regularization for sparse multi-task regression*, in "AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics", Naha, Japan, April 2019, vol. 89, https://hal.inria.fr/hal-02304176

[35] Z. RAMZI, P. CIUCIU, J.-L. STARCK.*Benchmarking proximal methods acceleration enhancements for CS-acquired MR image analysis reconstruction*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations Workshop", Toulouse, France, July 2019, https://hal.inria.fr/hal-02298569

[36] Z. RAMZI, P. CIUCIU, J.-L. STARCK.*Benchmarking Deep Nets MRI Reconstruction Models on the FastMRI Publicly Available Dataset*, in "ISBI 2020 - International Symposium on Biomedical Imaging", Iowa City, United States, April 2020, https://hal.inria.fr/hal-02436223

[37] M. SCETBON, G. VAROQUAUX.*Comparing distributions: $l1$ geometry improves kernel two-sample testing*, in "NeurIPS 2019 - 33th Conference on Neural Information Processing Systems", Vancouver, Canada, 2019, https://hal.inria.fr/hal-02292545

Conferences without Proceedings

[38] P. ABLIN, D. FAGOT, H. WENDT, A. GRAMFORT, C. FÉVOTTE.*A quasi-Newton algorithm on the orthogonal manifold for NMF with transform learning*, in "IEEE-ICASSP 2019 - International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, May 2019, https://arxiv.org/abs/1811.02225 , https://hal.archives-ouvertes.fr/hal-01912918

[39] P. ABLIN, A. GRAMFORT, J.-F. CARDOSO, F. BACH.*Stochastic algorithms with descent guarantees for ICA*, in "AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics", Naha, Japan, April 2019, https://hal.archives-ouvertes.fr/hal-02372092

[40] T. BAZEILLE, H. RICHARD, H. JANATI, B. THIRION.*Local Optimal Transport for Functional Brain Template Estimation*, in "IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging",

Hong Kong, China, June 2019 [*DOI :* 10.1007/978-3-030-20351-1_18], https://hal.archives-ouvertes.fr/hal-02278663

[41] H. CHERKAOUI, T. MOREAU, A. HALIMI, P. CIUCIU.*fMRI BOLD signal decomposition using a multivariate low-rank model*, in "Eusipco 2019 - 27th European Signal Processing Conference", Corunna, Spain, September 2019, https://hal.archives-ouvertes.fr/hal-02163497

[42] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*Online compressed sensing MR image reconstruction for high resolution T2* imaging*, in "ISMRM 2019 - 27th Annual Meeting and Exhibition", Montréal, Canada, May 2019, https://hal.archives-ouvertes.fr/hal-02314904

[43] L. EL GUEDDARI, P. CIUCIU, E. CHOUZENOUX, A. VIGNAUD, J.-C. PESQUET.*OSCAR-based reconstruction for compressed sensing and parallel MR imaging*, in "ISMRM 2019 - 27th Annual Meeting and Exhibition", Montréal, Canada, May 2019, https://hal.archives-ouvertes.fr/hal-02314911

[44] C. LAZARUS, P. WEISS, F. MAUCONDUIT, A. VIGNAUD, P. CIUCIU.*3D SPARKLING for accelerated ex vivo T2*-weighted MRI with compressed sensing*, in "ISMRM 2019 - 27th Annual Meeting & Exhibition", Montréal, France, May 2019, https://hal.archives-ouvertes.fr/hal-02374538

[45] M. MASSIAS, S. VAITER, A. GRAMFORT, J. SALMON.*Exploiting regularity in sparse Generalized Linear Models*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, https://hal-univ-bourgogne.archives-ouvertes.fr/hal-02288859

[46] T.-B. NGUYEN, J.-A. CHEVALIER, B. THIRION.*ECKO: Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data*, in "IPMI 2019 - International Conference on Information Processing in Medical Imaging", Hong Kong, Hong Kong SAR China, June 2019, https://hal.archives-ouvertes.fr/hal-02076510

[47] D. SABBAGH, P. ABLIN, G. VAROQUAUX, A. GRAMFORT, D. ENGEMANN.*Manifold-regression to predict from MEG/EEG brain signals without source modeling*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, https://hal.archives-ouvertes.fr/hal-02147708

Scientific Books (or Scientific Book chapters)

[48] L. ESCH, C. DINH, E. LARSON, D. ENGEMANN, M. JAS, S. KHAN, A. GRAMFORT, M. HÄMÄLÄINEN.*MNE: Software for Acquiring, Processing, and Visualizing MEG/EEG Data*, in "Magnetoencephalography", Springer International Publishing, October 2019, p. 355-371 [*DOI :* 10.1007/978-3-030-00087-5_59], https://hal.inria.fr/hal-02369299

Research Reports

[49] P. CIUCIU, A. KAZEYKINA.*Anisotropic compressed sensing for non-Cartesian MRI acquisitions*, CEA Paris Saclay ; Inria ; Université Paris Sud, Université Paris Saclay, October 2019, https://arxiv.org/abs/1910.14513 , https://hal.archives-ouvertes.fr/hal-02339689

Other Publications

[50] S. Abboud, D. Engemann, L. Cohen. *Semantic coding in the occipital cortex of early blind individuals*, February 2019, working paper or preprint [*DOI :* 10.1101/539437], https://hal.archives-ouvertes.fr/hal-02018272

[51] P. Ablin, T. Moreau, M. Massias, A. Gramfort. *Learning step sizes for unfolded sparse coding*, May 2019, working paper or preprint, https://hal.inria.fr/hal-02140383

[52] F. Baillin, A. Lefebvre, A. Pedoux, Y. Beauxis, D. Engemann, A. Maruani, F. Amsellem, T. Bourgeron, R. Delorme, G. Dumas. *Interactive Psychometrics for Autism with the Human Dynamic Clamp: Interpersonal Synchrony from Sensory-motor to Socio-cognitive Domains*, January 2020, Literature reference for the preprint first posted on medRxiv: https://www.medrxiv.org/content/10.1101/19013771v1 [*DOI :* 10.1101/19013771], https://hal.inria.fr/hal-02396923

[53] H. Banville, I. Albuquerque, A. Hyvärinen, G. Moffat, D. Engemann, A. Gramfort. *Self-supervised representation learning from electroencephalography signals*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02361350

[54] Q. Bertrand, M. Massias, A. Gramfort, J. Salmon. *Handling correlated and repeated measurements with the smoothed Multivariate square-root Lasso*, September 2019, https://arxiv.org/abs/1902.02509 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02010014

[55] P. Cerda, G. Varoquaux. *Encoding high-cardinality string categorical variables*, July 2019, https://arxiv.org/abs/1907.01860 - working paper or preprint, https://hal.inria.fr/hal-02171256

[56] L. El Gueddari, E. Chouzenoux, A. Vignaud, P. Ciuciu. *Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02292372

[57] D. Engemann, O. Kozynets, D. Sabbagh, G. Lemaître, G. Varoquaux, F. Liem, A. Gramfort. *Combining electrophysiology with MRI enhances learning of surrogate-biomarkers*, December 2019, working paper or preprint [*DOI :* 10.1101/856336], https://hal.archives-ouvertes.fr/hal-02395406

[58] B. Hermann, F. Raimondo, L. Hirsch, Y. Huang, M. Valente, P. Pérez, D. Engemann, F. Faugeras, N. Weiss, S. Demeret, B. Rohaut, L. Parra, J. D. Sitt, L. Naccache. *Combined behavioral and electrophysiological evidence for a direct cortical effect of prefrontal tDCS on disorders of consciousness*, April 2019, working paper or preprint [*DOI :* 10.1101/612309], https://hal.archives-ouvertes.fr/hal-02107872

[59] A. Jaiswal, J. Nenonen, M. Stenroos, A. Gramfort, S. S. Dalal, B. U. Westner, V. Litvak, J. C. Mosher, J.-M. Schoffelen, C. Witton, R. Oostenveld, L. Parkkonen. *Comparison of beamformer implementations for MEG source localization*, November 2019, working paper or preprint [*DOI :* 10.1101/795799], https://hal.inria.fr/hal-02369296

[60] H. Janati, T. Bazeille, B. Thirion, M. Cuturi, A. Gramfort. *Multi-subject MEG/EEG source imaging with sparse multi-task regression*, October 2019, working paper or preprint, https://hal.inria.fr/hal-02304194

[61] H. Janati, M. Cuturi, A. Gramfort. *Spatio-Temporal Alignments: Optimal transport through space and time*, October 2019, working paper or preprint, https://hal.inria.fr/hal-02309340

[62] J. JOSSE, N. PROST, E. SCORNET, G. VAROQUAUX. *On the consistency of supervised learning with missing values*, March 2019, https://arxiv.org/abs/1902.06931 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02024202

[63] D. LA ROCCA, P. CIUCIU, D. ENGEMANN, V. VAN WASSENHOVE. *Emergence of β and γ networks following multisensory training*, February 2019, working paper or preprint [*DOI :* 10.1101/560235], https://hal.archives-ouvertes.fr/hal-02052443

[64] C. LAZARUS, P. WEISS, L. E. GUEDDARI, F. MAUCONDUIT, A. VIGNAUD, P. CIUCIU. *3D SPARKLING trajectories for high-resolution T2*-weighted Magnetic Resonance imaging*, March 2019, working paper or preprint, https://hal.inria.fr/hal-02067080

[65] A. MACHLOUZARIDES-SHALIT, V. IOVENE, N. MAKRIS, D. WASSERMANN. *A Novel Sulcal Hierarchy Based on Manually Labelled Sulci*, June 2019, Organization for Human Brain Mapping, Poster, https://hal.inria.fr/hal-02347268

[66] M. MASSIAS, Q. BERTRAND, A. GRAMFORT, J. SALMON. *Support recovery and sup-norm convergence rates for sparse pivotal estimation*, January 2020, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02444978

[67] M. MASSIAS, S. VAITER, A. GRAMFORT, J. SALMON. *Dual Extrapolation for Sparse Generalized Linear Models*, August 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02263500

[68] V. MENON, G. GUILLERMO, M. PINSK, V.-D. NGUYEN, J.-R. LI, W. CAI, D. WASSERMANN. *Quantitative modeling links in vivo microstructural and macrofunctional organization of human and macaque insular cortex, and predicts cognitive control abilities*, January 2020, working paper or preprint [*DOI :* 10.1101/662601], https://hal.inria.fr/hal-02434382

[69] T. MOREAU, A. GRAMFORT. *DiCoDiLe: Distributed Convolutional Dictionary Learning*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02371715

[70] D. SABBAGH, P. ABLIN, G. VAROQUAUX, A. GRAMFORT, D. ENGEMANN. *Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02367411

[71] M.-A. SCHULZ, B. T. YEO, J. T. VOGELSTEIN, J. MOURAO-MIRANADA, J. N. KATHER, K. KORDING, B. A. RICHARDS, D. BZDOK. *Deep learning for brains?: Different linear and nonlinear scaling in UK Biobank brain images vs. machine-learning datasets*, September 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02276649

References in notes

[72] K. S. BUTTON, J. P. IOANNIDIS, C. MOKRYSZ, B. A. NOSEK, J. FLINT, E. S. ROBINSON, M. R. MUNAFÒ. *Power failure: why small sample size undermines the reliability of neuroscience*, in "Nature Reviews Neuroscience", 2013, vol. 14, n^{o} 5, p. 365–376

[73] R. A. POLDRACK, C. I. BAKER, J. DURNEZ, K. J. GORGOLEWSKI, P. M. MATTHEWS, M. R. MUNAFÒ, T. E. NICHOLS, J.-B. POLINE, E. VUL, T. YARKONI. *Scanning the horizon: towards transparent and reproducible neuroimaging research*, in "Nature Reviews Neuroscience", 2017, vol. 18, n^{o} 2, p. 115–126

Project-Team PARSIFAL

Proof search and reasoning with logic specifications

IN COLLABORATION WITH: Laboratoire d'informatique de l'école polytechnique (LIX)

IN PARTNERSHIP WITH:
CNRS

Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Proofs and Verification

Table of contents

Project-Team PARSIFAL

Creation of the Project-Team: 2007 July 01, end of the Project-Team: 2019 November 30

Keywords:

Computer Science and Digital Science:
A2.1. - Programming Languages
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.5. - Constraint programming
A2.1.10. - Domain-specific languages
A2.2.1. - Static analysis
A2.4.3. - Proofs
A2.5.4. - Software Maintenance & Evolution
A7.2.1. - Decision procedures
A7.2.2. - Automated Theorem Proving
A7.2.3. - Interactive Theorem Proving
A7.3.1. - Computational models and calculability
A9.8. - Reasoning

Other Research Topics and Application Domains:
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.8. - Reproducibility

1. Team, Visitors, External Collaborators

Research Scientists
Dale Miller [Team leader, Inria, Senior Researcher, HDR]
Beniamino Accattoli [Inria, Researcher]
Kaustuv Chaudhuri [Inria, Researcher]
François Lamarche [Inria, Senior Researcher]
Stéphane Lengrand [Researcher, until Dec 2019, HDR]
Ian Mackie [CNRS, Researcher]
Gabriel Scherer [Inria, Researcher]
Lutz Strassburger [Inria, Researcher, HDR]

External Collaborator
Andrea Condoluci [Università di Bologna, until Jul 2019]

Technical Staff
Matteo Acclavio [Telecom ParisTech, Engineer, from Dec 2019]

PhD Students
Ulysse Gerard [Université Paris-Diderot (Paris 7), PhD Student, until Oct 2019]
Maico Carlos Leberle [Inria, PhD Student]
Matteo Manighetti [Inria, PhD Student]
Emilie Grienenberger [Inria, PhD Student, until Oct 2022]
Marianela Morales [Inria, PhD Student, from Aug 2019]

Post-Doctoral Fellows

Marianna Girlando [Inria, Post-Doctoral Fellow, from Nov 2019]
Luc Pellissier [Inria, Post-Doctoral Fellow, from Nov 2019]
Benjamin Ralph [Inria, Post-Doctoral Fellow, from Feb 2019]

Visiting Scientist

Claudio Sacerdoti Coen [Università di Bologna, from Aug 2019 until Sep 2019]

Administrative Assistant

Maeva Jeannot [Inria, Administrative Assistant, until Oct 2019]

2. Overall Objectives

2.1. Main themes

The aim of the Parsifal team is to develop and exploit *proof theory* and *type theory* in the specification, verification, and analysis of computational systems.

- *Expertise*: the team conducts basic research in proof theory and type theory. In particular, the team is developing results that help with automated deduction and with the manipulation and communication of formal proofs.
- *Design*: based on experience with computational systems and theoretical results, the team develops new logical principles, new proof systems, and new theorem proving environments.
- *Implementation*: the team builds prototype systems to help validate basic research results.
- *Examples*: the design and implementation efforts are guided by examples of specification and verification problems. These examples not only test the success of the tools but also drive investigations into new principles and new areas of proof theory and type theory.

The foundational work of the team focuses on *structural* and *analytic* proof theory, *i.e.*, the study of formal proofs as algebraic and combinatorial structures and the study of proof systems as deductive and computational formalisms. The main focus in recent years has been the study of the *sequent calculus* and of the *deep inference* formalisms.

An important research question is how to reason about computational specifications that are written in a *relational* style. To this end, the team has been developing new approaches to dealing with induction, co-induction, and generic quantification. A second important question is of *canonicity* in deductive systems, *i.e.*, when are two derivations "essentially the same"? This crucial question is important not only for proof search, because it gives an insight into the structure and an ability to manipulate the proof search space, but also for the communication of *proof objects* between different reasoning agents such as automated theorem provers and proof checkers.

Important application areas currently include:

- Meta-theoretic reasoning on functional programs, such as terms in the λ-calculus
- Reasoning about behaviors in systems with concurrency and communication, such as the π-calculus, game semantics, *etc.*
- Combining interactive and automated reasoning methods for induction and co-induction
- Verification of distributed, reactive, and real-time algorithms that are often specified using modal and temporal logics
- Representing proofs as documents that can be printed, communicated, and checked by a wide range of computational logic systems.
- Development of cost models for the evaluation of proofs and programs.

3. Research Program

3.1. General overview

There are two broad approaches for computational specifications. In the *computation as model* approach, computations are encoded as mathematical structures containing nodes, transitions, and state. Logic is used to *describe* these structures, that is, the computations are used as models for logical expressions. Intensional operators, such as the modals of temporal and dynamic logics or the triples of Hoare logic, are often employed to express propositions about the change in state.

The *computation as deduction* approach, in contrast, expresses computations logically, using formulas, terms, types, and proofs as computational elements. Unlike the model approach, general logical apparatus such as cut-elimination or automated deduction becomes directly applicable as tools for defining, analyzing, and animating computations. Indeed, we can identify two main aspects of logical specifications that have been very fruitful:

- *Proof normalization*, which treats the state of a computation as a proof term and computation as normalization of the proof terms. General reduction principles such as β-reduction or cut-elimination are merely particular forms of proof normalization. Functional programming is based on normalization [51], and normalization in different logics can justify the design of new and different functional programming languages [32].
- *Proof search*, which views the state of a computation as a a structured collection of formulas, known as a *sequent*, and proof search in a suitable sequent calculus as encoding the dynamics of the computation. Logic programming is based on proof search [55], and different proof search strategies can be used to justify the design of new and different logic programming languages [54].

While the distinction between these two aspects is somewhat informal, it helps to identify and classify different concerns that arise in computational semantics. For instance, confluence and termination of reductions are crucial considerations for normalization, while unification and strategies are important for search. A key challenge of computational logic is to find means of uniting or reorganizing these apparently disjoint concerns.

An important organizational principle is structural proof theory, that is, the study of proofs as syntactic, algebraic and combinatorial objects. Formal proofs often have equivalences in their syntactic representations, leading to an important research question about *canonicity* in proofs – when are two proofs "essentially the same?" The syntactic equivalences can be used to derive normal forms for proofs that illuminate not only the proofs of a given formula, but also its entire proof search space. The celebrated *focusing* theorem of Andreoli [34] identifies one such normal form for derivations in the sequent calculus that has many important consequences both for search and for computation. The combinatorial structure of proofs can be further explored with the use of *deep inference*; in particular, deep inference allows access to simple and manifestly correct cut-elimination procedures with precise complexity bounds.

Type theory is another important organizational principle, but most popular type systems are generally designed for either search or for normalization. To give some examples, the Coq system [60] that implements the Calculus of Inductive Constructions (CIC) is designed to facilitate the expression of computational features of proofs directly as executable functional programs, but general proof search techniques for Coq are rather primitive. In contrast, the Twelf system [57] that is based on the LF type theory (a subsystem of the CIC), is based on relational specifications in canonical form (*i.e.*, without redexes) for which there are sophisticated automated reasoning systems such as meta-theoretic analysis tools, logic programming engines, and inductive theorem provers. In recent years, there has been a push towards combining search and normalization in the same type-theoretic framework. The Beluga system [58], for example, is an extension of the LF type theory with a purely computational meta-framework where operations on inductively defined LF objects can be expressed as functional programs.

The Parsifal team investigates both the search and the normalization aspects of computational specifications using the concepts, results, and insights from proof theory and type theory.

3.2. Inductive and co-inductive reasoning

The team has spent a number of years in designing a strong new logic that can be used to reason (inductively and co-inductively) on syntactic expressions containing bindings. This work is based on earlier work by McDowell, Miller, and Tiu [53] [52] [56] [61], and on more recent work by Gacek, Miller, and Nadathur [41] [40]. The Parsifal team, along with our colleagues in Minneapolis, Canberra, Singapore, and Cachan, have been building two tools that exploit the novel features of this logic. These two systems are the following.

- Abella, which is an interactive theorem prover for the full logic.
- Bedwyr, which is a model checker for the "finite" part of the logic.

We have used these systems to provide formalize reasoning of a number of complex formal systems, ranging from programming languages to the λ-calculus and π-calculus.

Since 2014, the Abella system has been extended with a number of new features. A number of new significant examples have been implemented in Abella and an extensive tutorial for it has been written [1].

3.3. Developing a foundational approach to defining proof evidence

The team is developing a framework for defining the semantics of proof evidence. With this framework, implementers of theorem provers can output proof evidence in a format of their choice: they will only need to be able to formally define that evidence's semantics. With such semantics provided, proof checkers can then check alleged proofs for correctness. Thus, anyone who needs to trust proofs from various provers can put their energies into designing trustworthy checkers that can execute the semantic specification.

In order to provide our framework with the flexibility that this ambitious plan requires, we have based our design on the most recent advances within the theory of proofs. For a number of years, various team members have been contributing to the design and theory of *focused proof systems* [35] [37] [38] [39] [43] [49] [50] and we have adopted such proof systems as the corner stone for our framework.

We have also been working for a number of years on the implementation of computational logic systems, involving, for example, both unification and backtracking search. As a result, we are also building an early and reference implementation of our semantic definitions.

3.4. Deep inference

Deep inference [44], [46] is a novel methodology for presenting deductive systems. Unlike traditional formalisms like the sequent calculus, it allows rewriting of formulas deep inside arbitrary contexts. The new freedom for designing inference rules creates a richer proof theory. For example, for systems using deep inference, we have a greater variety of normal forms for proofs than in sequent calculus or natural deduction systems. Another advantage of deep inference systems is the close relationship to category-theoretic proof theory. Due to the deep inference design one can directly read off the morphism from the derivations. There is no need for a counter-intuitive translation.

The following research problems are investigated by members of the Parsifal team:

- Find deep inference system for richer logics. This is necessary for making the proof theoretic results of deep inference accessible to applications as they are described in the previous sections of this report.
- Investigate the possibility of focusing proofs in deep inference. As described before, focusing is a way to reduce the non-determinism in proof search. However, it is well investigated only for the sequent calculus. In order to apply deep inference in proof search, we need to develop a theory of focusing for deep inference.

3.5. Proof nets, atomic flows, and combinatorial proofs

Proof nets graph-like presentations of sequent calculus proofs such that all "trivial rule permutations" are quotiented away. Ideally the notion of proof net should be independent from any syntactic formalism, but most notions of proof nets proposed in the past were formulated in terms of their relation to the sequent calculus. Consequently we could observe features like "boxes" and explicit "contraction links". The latter appeared not only in Girard's proof nets [42] for linear logic but also in Robinson's proof nets [59] for classical logic. In this kind of proof nets every link in the net corresponds to a rule application in the sequent calculus.

Only recently, due to the rise of deep inference, new kinds of proof nets have been introduced that take the formula trees of the conclusions and add additional "flow-graph" information (see e.g., [48][2] leading to the notion of *atomic flow* and [45]. On one side, this gives new insights in the essence of proofs and their normalization. But on the other side, all the known correctness criteria are no longer available.

Combinatorial proofs [47] are another form syntax-independent proof presentation which separates the multiplicative from the additive behaviour of classical connectives.

The following research questions investigated by members of the Parsifal team:

- Finding (for classical and intuitionistic logic) a notion of canonical proof presentation that is deductive, i.e., can effectively be used for doing proof search.
- Studying the normalization of proofs using atomic flows and combinatorial proofs, as they simplify the normalization procedure for proofs in deep inference, and additionally allow to get new insights in the complexity of the normalization.
- Studying the size of proofs in the combinatorial proof formalism.

3.6. Cost Models and Abstract Machines for Functional Programs

In the *proof normalization* approach, computation is usually reformulated as the evaluation of functional programs, expressed as terms in a variation over the λ-calculus. Thanks to its higher-order nature, this approach provides very concise and abstract specifications. Its strength is however also its weakness: the abstraction from physical machines is pushed to a level where it is no longer clear how to measure the complexity of an algorithm.

Models like Turing machines or RAM rely on atomic computational steps and thus admit quite obvious cost models for time and space. The λ-calculus instead relies on a single non-atomic operation, β-reduction, for which costs in terms of time and space are far from evident.

Nonetheless, it turns out that the number of β-steps is a reasonable time cost model, i.e.,it is polynomially related to those of Turing machines and RAM. For the special case of *weak evaluation* (i.e., reducing only β-steps that are not under abstractions)—which is used to model functional programming languages—this is a relatively old result due to Blelloch and Greiner [36] (1995). It is only very recently (2014) that the strong case—used in the implementation models of proof assistants—has been solved by Accattoli and Dal Lago [33].

With the recent recruitment of Accattoli, the team's research has expanded in this direction. The topics under investigations are:

1. *Complexity of Abstract Machines*. Bounding and comparing the overhead of different abstract machines for different evaluation schemas (weak/strong call-by-name/value/need λ-calculi) with respect to the cost model. The aim is the development of a complexity-aware theory of the implementation of functional programs.
2. *Reasonable Space Cost Models*. Essentially nothing is known about reasonable space cost models. It is known, however, that environment-based execution model—which are the mainstream technology for functional programs—do not provide an answer. We are exploring the use of the non-standard implementation models provided by Girard's Geometry of Interaction to address this question.

4. Application Domains

4.1. Automated Theorem Proving

The Parsifal team studies the structure of mathematical proofs, in ways that often makes them more amenable to automated theorem proving – automatically searching the space of proof candidates for a statement to find an actual proof – or a counter-example.

(Due to fundamental computability limits, fully-automatic proving is only possible for simple statements, but this field has been making a lot of progress in recent years, and is in particular interested with the idea of generating verifiable evidence for the proofs that are found, which fits squarely within the expertise of Parsial.)

4.2. Proof-assistants

The team work on the structure of proofs also suggests ways that they could be presented to a user, edited and maintained, in particular in "proof assistants", automated tool to assist the writing of mathematical proofs with automatic checking of their correctness.

4.3. Programming language design

Our work also gives insight on the structure and properties of programming languages. We can improve the design or implementation of programming languages, help programmers or language implementors reason about the correctness of the programs in a given language, or reason about the cost of execution of a program.

5. Highlights of the Year

5.1. Highlights of the Year

- The journal Mathematical Structures in Computer Science published "A special issue on structural proof theory, automated reasoning and computation in celebration of Dale Miller's 60th birthday" – volume 29, Special issue 8, September 2019.
- Accattoli was invited speaker at the international conference FSCD 2019.

6. New Software and Platforms

6.1. Abella

FUNCTIONAL DESCRIPTION: Abella is an interactive theorem prover for reasoning about computations given as relational specifications. Abella is particuarly well suited for reasoning about binding constructs.

- Participants: Dale Miller, Gopalan Nadathur, Kaustuv Chaudhuri, Mary Southern, Matteo Cimini, Olivier Savary-Bélanger and Yuting Wang
- Partner: Department of Computer Science and Engineering, University of Minnesota
- Contact: Kaustuv Chaudhuri
- URL: http://abella-prover.org/

6.2. Bedwyr

Bedwyr - A proof search approach to model checking

KEYWORD: Model Checker

FUNCTIONAL DESCRIPTION: Bedwyr is a generalization of logic programming that allows model checking directly on syntactic expressions that possibly contain bindings. This system, written in OCaml, is a direct implementation of two recent advances in the theory of proof search.

It is possible to capture both finite success and finite failure in a sequent calculus. Proof search in such a proof system can capture both may and must behavior in operational semantics. Higher-order abstract syntax is directly supported using term-level lambda-binders, the nabla quantifier, higher-order pattern unification, and explicit substitutions. These features allow reasoning directly on expressions containing bound variables.

The distributed system comes with several example applications, including the finite pi-calculus (operational semantics, bisimulation, trace analyses, and modal logics), the spi-calculus (operational semantics), value-passing CCS, the lambda-calculus, winning strategies for games, and various other model checking problems.

- Participants: Dale Miller, Quentin Heath and Roberto Blanco Martinez
- Contact: Dale Miller
- URL: http://slimmer.gforge.inria.fr/bedwyr/

6.3. Checkers

Checkers - A proof verifier

KEYWORDS: Proof - Certification - Verification

FUNCTIONAL DESCRIPTION: Checkers is a tool in Lambda-prolog for the certification of proofs. Checkers consists of a kernel which is based on LKF and is based on the notion of ProofCert.

- Participants: Giselle Machado Nogueira Reis, Marco Volpe and Tomer Libal
- Contact: Tomer Libal
- URL: https://github.com/proofcert/checkers

6.4. Psyche

Proof-Search factorY for Collaborative HEuristics

KEYWORD: Automated theorem proving

FUNCTIONAL DESCRIPTION: Psyche is a modular platform for automated or interactive theorem proving, programmed in OCaml and built on an architecture (similar to LCF) where a trusted kernel interacts with plugins. The kernel offers an API of proof-search primitives, and plugins are programmed on top of the API to implement search strategies. This architecture is set up for pure logical reasoning as well as for theory-specific reasoning, for various theories.

RELEASE FUNCTIONAL DESCRIPTION: It is now equipped with the machinery to handle quantifiers and quantifier-handling techniques. Concretely, it uses meta-variables to delay the instantiation of existential variables, and constraints on meta-variables are propagated through the various branches of the search-space, in a way that allows local backtracking. The kernel, of about 800 l.o.c., is purely functional.

- Participants: Assia Mahboubi, Jean-Marc Notin and Stéphane Graham-Lengrand
- Contact: Stéphane Graham-Lengrand
- URL: http://www.csl.sri.com/users/sgl/

6.5. Maetning

FUNCTIONAL DESCRIPTION: Mætning is an automated theorem prover for intuitionistic predicate logic that is designed to disprove non-theorems.

- Contact: Kaustuv Chaudhuri
- URL: https://github.com/chaudhuri/maetning/

6.6. OCaml

KEYWORDS: Functional programming - Static typing - Compilation

FUNCTIONAL DESCRIPTION: The OCaml language is a functional programming language that combines safety with expressiveness through the use of a precise and flexible type system with automatic type inference. The OCaml system is a comprehensive implementation of this language, featuring two compilers (a bytecode compiler, for fast prototyping and interactive use, and a native-code compiler producing efficient machine code for x86, ARM, PowerPC and System Z), a debugger, a documentation generator, a compilation manager, a package manager, and many libraries contributed by the user community.

- Participants: Damien Doligez, Xavier Leroy, Fabrice Le Fessant, Luc Maranget, Gabriel Scherer, Alain Frisch, Jacques Garrigue, Marc Shinwell, Jeremy Yallop and Leo White
- Contact: Damien Doligez
- URL: https://ocaml.org/

7. New Results

7.1. Functional programming with λ-tree syntax

Participants: Ulysse Gerard, Dale Miller, Gabriel Scherer.

We have been designing a new functional programming language, MLTS, that uses the λ-*tree* syntax approach to encoding bindings that appear within data structures [20]. In this setting, bindings never become free nor escape their scope: instead, binders in data structures are permitted to *move* into binders within programs phrases. The design of MLTS—whose concrete syntax is based on that of OCaml—includes additional sites within programs that directly support this movement of bindings. Our description of MLTS includes a typing discipline that naturally extends the typing of OCaml programs.

In addition to the natural semantics for MLTS that we proposed in 2018, we also have a small-step operational semantics which gives in particular a fine-grained description of the runtime behavior of the ∇ operator in patterns. It leads in particular to a direct implementation in Lambda-Prolog (which does not contain a native ∇ operator) that allows more expressive constructs (higher-arity types) than our previous presentation.

7.2. Mechanized metatheory revisited

Participant: Dale Miller.

When proof assistants and theorem provers implement the metatheory of logical systems, they must deal with a range of syntactic expressions (e.g., types, formulas, and proofs) that involve variable bindings. Since most mature proof assistants do not have built-in methods to treat bindings, they have been extended with various packages and libraries that allow them to encode such syntax using, for example, De Bruijn numerals. In the paper, [10], Miller puts forward the argument that bindings are such an intimate aspect of the structure of expressions that they should be accounted for directly in the underlying programming language support for proof assistants and not via packages and libraries. He presents an approach to designing programming languages and proof assistants that directly supports bindings in syntax. The roots of this approach can be found in the *mobility* of binders between term-level bindings, formula-level bindings (quantifiers), and proof-level bindings (eigenvariables). In particular, the combination of Church's approach to terms and formulas (found in his Simple Theory of Types) and Gentzen's approach to proofs (found in his sequent calculus) yields a framework for the interaction of bindings with a full range of logical connectives and quantifiers. Miller also illustrates how that framework provides a direct and semantically clean treatment of computation and reasoning with syntax containing bindings.

7.3. New applications of Foundational Proof Certificates

Participants: Kaustuv Chaudhuri, Matteo Manighetti, Dale Miller.

The formal framework of *Foundational Proof Certificates* (FPC) was developed in previous years within the Parsifal team. We continue to push on their applications in a number of settings in computational logic. In 2019, we developed two such new applications.

In order to apply FPCs to the conventional setting of classical logic theorem provers, the FPC setting needed to treat proof evidence containing Skolem functions. Using FPC directly meant that we needed to do such certification without using the mathematical concepts of model-theoretic semantics (i.e., preservation of satisfiability) and choice principles (i.e., epsilon terms). Instead, our proof checking kernel is an implementation of Gentzen's sequent calculus, which directly supports quantifier alternation by using eigenvariables. In [19], we described deskolemization as a mapping from client-side terms, used in proofs generated by theorem provers, into kernel-side terms, used within our proof checking kernel. This mapping which associates skolemized terms to eigenvariables relies on using outer skolemization.

Property-based testing (PBT) is a technique for validating code against an executable specification by automatically generating test-data. In the paper [18], we presented a proof-theoretical reconstruction of this style of testing for relational specifications and employ FPCs to describe test generators. We did this by presenting certain kinds of "proof outlines" that can be used to describe various common generation strategies in the PBT literature, ranging from random to exhaustive, including their combination. We also address the shrinking of counterexamples as a first step towards their explanation. Once generation is accomplished, the testing phase boils down to a standard logic programming search. We could also we lift our techniques to treat data structures containing bindings using λ-*tree* syntax. The λProlog programming language is capable of performing both the generation and checking of tests. We validated this approach by tackling benchmarks in the metatheory of programming languages coming from related tools such as PLT-Redex Property-Based Testing via Proof Reconstruction. This work was done in collaboration with Roberto Blanco, a postdoc from Inria Paris, and Alberto Momigliano, a professor from the University of Milan.

7.4. Historical reflections on proof theory and logic programming

Participant: Dale Miller.

Miller has been working in the area of logic programming and proof theory for more than three decades. Some of his historical reflections on how these two topics influenced each other are contained in the paper [11]. While it is widely known that proof theory has been helpful in shaping the development of logic programming, particular of extensions to conventional Prolog, this paper also documents a few specific examples where logic programming influenced the development of some topics in proof theory.

7.5. Intuitionistic proofs without syntax

Participant: Lutz Straßburger.

We present Intuitionistic Combinatorial Proofs (ICPs), a concrete geometric semantics of intuitionistic logic based on the principles of classical combinatorial proofs. An ICP naturally factorizes into a linear fragment, a graphical abstraction of an IMLL proof net (an arena net), and a parallel contraction-weakening fragment (a skew fibration). ICPs relate to game semantics, and can be seen as a strategy in a Hyland-Ong arena, generalized from a tree-like to a dag-like strategy. Our first main result, Polynomial Full Completeness, is thatICPs as a semantics are complexity-aware: the translations to and from sequent calculus are size-preserving (up to a polynomial). By contrast, lambda-calculus and game semantics incur an exponential blowup. Our second main result, Local Canonicity, is that ICPs abstract fully and faithfully over the non-duplicating permutations of the sequent calculus. These results have been presented at the LICS 2019 conference [23].

7.6. Towards a combinatorial proof theory

Participants: Lutz Straßburger, Benjamin Ralph.

The main part of a classical combinatorial proof is a skew fibration, which precisely captures the behavior of weakening and contraction. Relaxing the presence of these two rules leads to certain substructural logics and substructural proof theory. We investigated what happens if we replace the skew fibration by other kinds of graph homomorphism. This leads us to new logics and proof systems that we call combinatorial. This has been presented at the TABLEAUX 2019 conference [22].

7.7. Combinatorial Proofs for Logics of Relevance and Entailment

Participants: Lutz Straßburger, Matteo Acclavio.

In this work (presented at the WoLLIC 2019 conference [16]) we characterize classical combinatorial proofs which also represent valid proofs for relevant logic with and without the mingle axiom. Moreover, we extend our syntax in order to represent combinatorial proofs for the more restrictive framework of entailment logic.

7.8. On combinatorial proofs for modal logic

Participants: Lutz Straßburger, Matteo Acclavio.

In this work [17], we extend combinatorial proofs to modal logics. The crucial ingredient for modeling the modalities is the use of a self-dual non-commutative operator that has first been observed by Retoré through pomset logic. Consequently, we had to generalize the notion of skew fibration from cographs to Guglielmi's relation webs. Our main result is a sound and complete system of combinatorial proofs for all normal and non-normal modal logics in the S4-tesseract. The proof of soundness and completeness is based on the sequent calculus with some added features from deep inference.

7.9. Deep inference and expansion trees for second-order multiplicative linear logic

Participant: Lutz Straßburger.

In this work, we introduce the notion of expansion tree for linear logic. As in Miller's original work, we have a shallow reading of an expansion tree that corresponds to the conclusion of the proof, and a deep reading which is a formula that can be proved by propositional rules. We focus our attention to MLL2, and we also present a deep inference system for that logic. This allows us to give a syntactic proof to a version of Herbrand's theorem. This has been published in an special issue of MSCS [12].

7.10. A fully labelled proof system for intuitionistic modal logics

Participants: Lutz Straßburger, Marianela Morales.

In this paper we present a labelled sequent system for intuitionistic modal logics such that there is not only one, but two relation symbols appearing in sequents: one for the accessibility relation associated with the Kripke semantics for normal modal logics and one for the preorder relation associated with the Kripke semantics for intuitionistic logic. This puts our system in close correspondence with the standard birelational Kripke semantics for intuitionistic modal logics. As a consequence it can encompass a wider range of intuitionistic modal logics than existing labelled systems. We also show an internal cut elimination proof for our system [30].

7.11. Types by Need

Participants: Beniamino Accattoli, Maico Leberle.

This joint work with Giulio Guerrieri (Post-doc at Bath University) [27] develops a multi type system for call-by-need evaluation of the -calculus. The type system is obtained by combining features by well-known systems for call-by-name and call-by-value. It characterizes termination, and, moreover, its type derivations provide precise information about the number of steps to reach the result. The novelty is that, while the systems for call-by-name and call-by-value are obtained by the linear logic interpretation of these evaluation schemes, call-by-need has no linear logic interpretation.

7.12. Sharing Equality is Linear

Participants: Beniamino Accattoli, Andrea Condoluci, Claudio Sacerdoti Coen.

This work [28] studies how to compare higher-order programs with sharing for sharing equality, that is, for equality of their unshared underlying programs. The point, of course, is to do it efficiently, without unsharing the programs, that would otherwise introduce an exponential blow-up. We develop the first algorithm linear in the size of the shared terms, by adapting the famous Patterson and Wegman algorithm for first-order unification.

7.13. Crumbling Abstract Machines

Participants: Beniamino Accattoli, Andrea Condoluci, Claudio Sacerdoti Coen.

This joint work with Giulio Guerrieri (Post-doc at Bath University) [26] studies a new compilation technique for functional programs, dubbed *crumbling* and resembling the transformation into administrative normal form of Flanagan, Sabry, Duba, and Felleisen. It is shown that it simplifies the design of abstract machines without altering the complexity of the overhead. Moreover, it smoothly scales up to open terms and it does not suffer of the slowdowns of administrative normal forms pointed out by Kennedy.

7.14. Factorization and Normalization, Essentially

Participant: Beniamino Accattoli.

This joint work with Claudia Faggian (CNRS researcher at Paris Diderot) and Giulio Guerrieri (Post-doc at Bath University) [15] refines a rewriting technique for proving factorization and normalization theorems for λ-calculi, that are theorems providing foundations to the design of functional programming languages and proof assistants. We both simplify and extend the scope of a widely used technique by Takahashi. At the concrete level, the new abstract technique is applied to four relevant case studies.

7.15. A Fresh Look at the -Calculus

Participant: Beniamino Accattoli.

This paper [25] is the trace of the invited talk given by Accattoli at FSCD 2019. More than just an abstract, the paper is a lengthy overview of the research on λ-calculus, cost models, sharing, and abstract machines pursued by Accattoli and his co-authors in the last 10 years.

7.16. Abstract Machines for Open Call-by-Value

Participant: Beniamino Accattoli.

This journal paper in collaboration with Giulio Guerrieri (Post-doc at Bath University) [4] outlines a theory of abstract machines for the call-by-value -calculus with open terms. It refines and extends the results by the same authors from 2017, which were among the selected ones from the international conference FSEN 2017 for publication in a journal.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. OCaml Software Foundation

Participant: Gabriel Scherer.

The OCaml Software Foundation (OCSF), [0] established in 2018 under the umbrella of the Inria Foundation, aims to promote, protect, and advance the OCaml programming language and its ecosystem, and to support and facilitate the growth of a diverse and international community of OCaml users.

[0] http://ocaml-sf.org/

Gabriel Scherer serves as the director of the foundation.

8.1.2. Funding from Nomadic Labs

Participant: Gabriel Scherer.

Nomadic Labs, a Paris-based company, has implemented the Tezos blockchain and cryptocurrency entirely in OCaml. This year, Nomadic Labs and Inria have signed a framework agreement (“contrat-cadre”) that allows Nomadic Labs to fund multiple research efforts carried out by Inria groups. Within this framework, we participate to two 3-year grants, in collaboration with the Cambium team at Inria Paris:

- “Évolution d’OCaml”. This grant is intended to fund a number of improvements to OCaml, including the addition of new features and a possible re-design of the OCaml type-checker.
- “Maintenance d’OCaml”. This grant is intended to fund the day-to-day maintenance of OCaml as well as the considerable work involved in managing the release cycle.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. DIM-RFSI

Gabriel Scherer obtained funding from the Région Île-de-France to hire a post-doc, Luc Pellissier, to work on canonical representation of programs (linking proof theory and category-theory approaches), in collaboration with Adrien Guatto in IRIF (Université Paris 7).

9.2. National Initiatives

9.2.1. ANR

COCA HOLA: Cost Models for Complexity Analyses of Higher-Order Languages, coordinated by B. Accattoli, 2016–2019.

FISP: The Fine Structure of Formal Proof Systems and their Computational Interpretations, coordinated by Lutz Straßburger in collaboration with Université Paris 7, Universität Innsbruck and TU Wien, 2016–2019.

9.2.2. Competitivity Clusters

UPScale: Universality of Proofs in SaCLay, a Working Group of LabEx DigiCosme, organized by Chantal Keller (LRI) with regular participation from Parsifal members and a post-doc co-supervision.

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Claudio Sacerdoti Coen (Universita di Bologna, Italy) spent a month visiting Beniamino Accattoli thanks to funding for short-term international visits.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

Dale Miller is currently General Chair for the ACM/IEEE Symposium on Logic in Computer Science (LICS).

10.1.1.2. Member of the Organizing Committees

Dale Miller is member of the Steering Committee for CPP, FSCD, and LFMTP and is a member of the Executive Committee of the ACM Special Interest Group on Logic and Compuation (SIGLOG).

Gabriel Scherer is part of the Steering Committee of the ML Family Workshop and the OCaml Workshop.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

Dale Miller was program committee co-chair for the workshop on Logical Frameworks and Meta Languages: Theory and Practice (LFMTP), 2019, Vancouver, Canada.

10.1.2.2. Member of the Conference Program Committees

- Gabriel Scherer: JFLA 2020.
- Beniamino Accattoli: LSFA 2019.
- Dale Miller is a PC member for the 10th International Joint Conference on Automated Reasoning (IJCAR-2020), the 23rd International Conference on Logic for Programming, Artificial Intelligence and Reasoning (LPAR-23), and the Workshop on Trends, Extensions, Applications and Semantics of Logic Programming (TEASE-LP).

10.1.2.3. Reviewer

- Gabriel Scherer: POPL 2020, ESOP 2020.
- Lutz Straßburger: FoSSaCS 2020, CSL 2020, TABLEAUX 2019, WoLLIC 2019, LICS 2019.
- Beniamino Accattoli: FoSSaCS 2020, POPL 2020, ICALP 2019, ICTAC 2019, PPDP 2019, FSCD 2019, CSL 2019, LSFA 2019.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Dale Miller is on the editorial board of the Journal of Automated Reasoning (Springer) and the Jounral of Applied Logic (Elsevier). He has also been an editor for a special issue of FSCD 2017 for Logical Methods in Computer Science.

10.1.3.2. Reviewer - Reviewing Activities

- Lutz Straßburger: APAL, ToCL, BSL, LMCS.
- Beniamino Accattoli: LMCS, TCS.
- Dale Miller has served on the evaluation committee for the EATCS Distinguished Dissertation Award and the EACSL Ackermann Award. He was also the Chair of the Herbrand Award Committee of the Association for Automated Reasoning.

10.1.4. Invited Talks

- Beniamino Accattoli has been invited speaker at FSCD 2019.
- Dale Miller has been an invited speaker at the Workshop on Proof Theory for Automated Deduction, Automated Deduction for Proof Theory (23-25 October 2019, Funchal, Madeira) and the Third Tübingen Conference on Proof-Theoretic Semantics, 27-30 March 2019.

10.1.5. Scientific Expertise

- Lutz Straßburger: reviewer for the Israel Science Foundation.
- Dale Miller has been a member of an international review panel for the Distinguished Professor Grant at the Swedish Research Council.
- Gabriel Scherer: reviewer for ANR.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : G. Scherer, "*Introduction á la programmation fonctionnelle*", 50h, L1, Paris 8 (Vincennes - Saint Denis), France

Master: B. Accattoli, "*Logique linéaire et paradigmes logiques du calcul*", 18h, M2, Master Parisien de Recherche en Informatique (MPRI), France.

Master: D. Miller, "*Logique linéaire et paradigmes logiques du calcul*", 18h, M2, Master Parisien de Recherche en Informatique (MPRI), France.

Summer School: G. Scherer, "*Programmation fonctionnelle en OCaml*", 12h, public d'ingénieurs de recherche, formation ANF (CNRS), France

Summer School: L. Straßburger: "*Introduction to Deep Inference*", 10h, ESSLLI 2020, Riga, Latvia

Summer School: B. Accattoli, "*λ-Calculus and Reasonable Cost Models*", 15h, Escuela de Ciencia Informaticas (ECI 2019), Buenos Aires, Argentina.

10.2.2. Supervision

- PhD in progress: Marianela Morales, "Combinatorial Proof Theory for Modal Logic", 1/10/2019, Lutz Straßburger.
- PhD in progress: Maico Leberle, "Call-by-need and Reasonable Cost Models", Beniamino Accattoli.
- PhD completed: Ulysse Gérard, "Computing with relations, functions, and bindings", 18 October 2019, Ecole Doctorale de l'Institut Polytechnique de Paris, advised by Dale Miller.
- PhD in progress: Matteo Manighetti, "Structural proof theory for induction in linear logic", advised by Dale Miller since 1/10/2017.
- PhD in progress: Marianela Morales, "Combinatorial Proof Theory for Modal Logic", 1/10/2019, Lutz Straßburger

10.2.3. Juries

Dale Miller was a reportor for the PhD thesis of Aurore Alcolei (ENS Lyon, 17 October 2019).

10.3. Popularization

10.3.1. Interventions

G. Scherer and M. Manighetti participated the "Fête de la Science" exhibit at Inria Saclay on the whole day of October 11th, 2019.

G. Scherer presented the research domain of certified programming to an audience of computer security professionals at the "Pass the Salt" conference in Lille, on Wednesday July 3rd.

11. Bibliography

Major publications by the team in recent years

[1] D. Baelde, K. Chaudhuri, A. Gacek, D. Miller, G. Nadathur, A. Tiu, Y. Wang.*Abella: A System for Reasoning about Relational Specifications*, in "Journal of Formalized Reasoning", 2014, vol. 7, n^o 2, p. 1-89 [*DOI :* 10.6092/ISSN.1972-5787/4650], https://hal.inria.fr/hal-01102709

[2] A. GUGLIELMI, T. GUNDERSEN, L. STRASSBURGER.*Breaking Paths in Atomic Flows for Classical Logic*, in "Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science (LICS 2010)", Edinburgh, United Kingdom, July 2010, p. 284–293 [*DOI :* 10.1109/LICS.2010.12], http://www.lix.polytechnique.fr/~lutz/papers/AFII.pdf

Publications of the year

Doctoral Dissertations and Habilitation Theses

[3] U. GÉRARD.*Computing with relations, functions, and bindings*, École Polytechnique ; Inria Saclay, October 2019, https://hal.archives-ouvertes.fr/tel-02414237

Articles in International Peer-Reviewed Journal

[4] B. ACCATTOLI, G. GUERRIERI.*Abstract Machines for Open Call-by-Value*, in "Science of Computer Programming", 2019, vol. 184 [*DOI :* 10.1016/J.SCICO.2019.03.002], https://hal.archives-ouvertes.fr/hal-02415780

[5] T. BROCK-NANNESTAD, D. ILIK.*An Intuitionistic Formula Hierarchy Based on High-School Identities*, in "Mathematical Logic Quarterly", May 2019, vol. 65, nO 1, p. 57-79, https://arxiv.org/abs/1601.04876 [*DOI :* 10.1002/MALQ.201700047], https://hal.inria.fr/hal-01354181

[6] K. CHAUDHURI, C. OLARTE, E. PIMENTEL, J. DESPEYROUX.*Hybrid Linear Logic, revisited*, in "Mathematical Structures in Computer Science", 2019, forthcoming [*DOI :* 10.1017/S0960129518000439], https://hal.inria.fr/hal-01968154

[7] J. COURTIEL, K. YEATS, N. ZEILBERGER.*Connected chord diagrams and bridgeless maps*, in "The Electronic Journal of Combinatorics", November 2019, https://hal.archives-ouvertes.fr/hal-01650141

[8] Q. HEATH, D. MILLER.*A proof theory for model checking*, in "Journal of Automated Reasoning", December 2019, vol. 63, nO 4, p. 857-885 [*DOI :* 10.1007/S10817-018-9475-3], https://hal.inria.fr/hal-01814006

[9] R. KUZNETS, L. STRASSBURGER.*Maehara-style modal nested calculi*, in "Archive for Mathematical Logic", May 2019, vol. 58, nO 3-4, p. 359-385 [*DOI :* 10.1007/S00153-018-0636-1], https://hal.inria.fr/hal-01942240

[10] D. MILLER.*Mechanized metatheory revisited*, in "Journal of Automated Reasoning", October 2019, vol. 63, nO 3, p. 625-665 [*DOI :* 10.1007/S10817-018-9483-3], https://hal.inria.fr/hal-01884210

[11] D. MILLER.*Reciprocal Influences Between Proof Theory and Logic Programming*, in "Philosophy & Technology", August 2019 [*DOI :* 10.1007/S13347-019-00370-X], https://hal.inria.fr/hal-02368867

[12] L. STRASSBURGER.*Deep inference and expansion trees for second-order multiplicative linear logic*, in "Mathematical Structures in Computer Science", September 2019, vol. 29, p. 1030-1060 [*DOI :* 10.1017/S0960129518000385], https://hal.inria.fr/hal-01942410

[13] L. STRASSBURGER.*On the decision problem for MELL*, in "Theoretical Computer Science", May 2019, vol. 768, p. 91-98 [*DOI :* 10.1016/J.TCS.2019.02.022], https://hal.inria.fr/hal-02386746

[14] L. STRASSBURGER.*The problem of proof identity, and why computer scientists should care about Hilbert's 24th problem*, in "Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences", January 2019, vol. 377, n$^{\text{o}}$ 2140, 20180038 [*DOI :* 10.1098/RSTA.2018.0038], https://hal.inria.fr/hal-02475417

International Conferences with Proceedings

[15] B. ACCATTOLI, C. FAGGIAN, G. GUERRIERI.*Factorization and Normalization, Essentially*, in "APLAS 2019 - 17th Asian Symposium on Programming Languages and Systems", Bali, Indonesia, Springer, December 2019 [*DOI :* 10.1007/978-3-030-34175-6_9], https://hal.archives-ouvertes.fr/hal-02411556

[16] M. ACCLAVIO, L. STRASSBURGER.*On Combinatorial Proofs for Logics of Relevance and Entailment*, in "WoLLIC 2019 - 26th International Workshop on Logic, Language, Information, and Computation", Utrecht, Netherlands, June 2019, p. 1-16 [*DOI :* 10.1007/978-3-662-59533-6_1], https://hal.inria.fr/hal-02390426

[17] M. ACCLAVIO, L. STRASSBURGER.*On Combinatorial Proofs for Modal Logic*, in "TABLEAUX 2019 - 28t International Conference on Automated Reasoning with Analytic Tableaux and Related Methods", London, United Kingdom, Springer, September 2019, p. 223-240 [*DOI :* 10.1007/978-3-030-29026-9_13], https://hal.inria.fr/hal-02390400

[18] R. BLANCO, D. MILLER, A. MOMIGLIANO.*Property-Based Testing via Proof Reconstruction*, in "PPDP 2019 - 21st International Symposium on Principles and Practice of Programming Languages", Porto, Portugal, ACM Press, October 2019, p. 1-13 [*DOI :* 10.1145/3354166.3354170], https://hal.inria.fr/hal-02368931

[19] K. CHAUDHURI, M. MANIGHETTI, D. MILLER.*A Proof-Theoretic Approach to Certifying Skolemization*, in "CPP 2019 - 8th ACM SIGPLAN International Conference", Cascais, Portugal, ACM Press, January 2019, p. 78-90 [*DOI :* 10.1145/3293880.3294094], https://hal.inria.fr/hal-02368946

[20] U. GÉRARD, D. MILLER, G. SCHERER.*Functional programming with λ-tree syntax*, in "PPDP 2019 - 21st International Symposium on Principles and Practice of Programming Languages", Porto, Portugal, ACM Press, October 2019, p. 1-16 [*DOI :* 10.1145/3354166.3354177], https://hal.inria.fr/hal-02368906

[21] W. HEIJLTJES, D. J. D. HUGHES, L. STRASSBURGER.*Proof Nets for First-Order Additive Linear Logic*, in "FSCD 2019 - 4th International Conference on Formal Structures for Computation and Deduction", Dortmund, Germany, Proceedings of 4th International Conference on Formal Structures for Computation and Deduction (FSCD 2019), June 2019, vol. 131, p. 22:1-22:22 [*DOI :* 10.4230/LIPICS.FSCD.2019.22], https://hal.inria.fr/hal-02386942

[22] B. RALPH, L. STRASSBURGER.*Towards a Combinatorial Proof Theory*, in "TABLEAUX 2019 - 28th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods", London, United Kingdom, August 2019, p. 259-276 [*DOI :* 10.1007/978-3-030-29026-9_15], https://hal.inria.fr/hal-02390417

[23] L. STRASSBURGER, W. HEIJLTJES, D. J. D. HUGHES.*Intuitionistic proofs without syntax*, in "LICS 2019 - 34th Annual ACM/IEEE Symposium on Logic in Computer Science", Vancouver, Canada, IEEE, June 2019, p. 1-13 [*DOI :* 10.1109/LICS.2019.8785827], https://hal.inria.fr/hal-02386878

National Conferences with Proceeding

[24] S. COLIN, R. LEPIGRE, G. SCHERER.*Unboxing Mutually Recursive Type Definitions in OCaml*, in "JFLA 2019 - 30 èmes journées francophones des langages applicatifs", Les Rousses, France, January 2019, https://arxiv.org/abs/1811.02300 , https://hal.inria.fr/hal-01929508

Conferences without Proceedings

[25] B. ACCATTOLI.*A Fresh Look at the λ-Calculus*, in "FSCD 2019 - 4th International Conference on Formal Structures for Computation and Deduction", Dortmund, Germany, June 2019 [*DOI :* 10.4230/LIPIcs.FSCD.2019.1], https://hal.archives-ouvertes.fr/hal-02415786

[26] B. ACCATTOLI, A. CONDOLUCI, G. GUERRIERI, C. S. COEN.*Crumbling Abstract Machines*, in "PPDP 2019 - 21st International Symposium on Principles and Practice of Programming Languages", Porto, Portugal, October 2019 [*DOI :* 10.1145/3354166.3354169], https://hal.archives-ouvertes.fr/hal-02415766

[27] B. ACCATTOLI, G. GUERRIERI, M. LEBERLE.*Types by Need*, in "ESOP 2019 - 28th European Symposium on Programming", Prague, Czech Republic, April 2019 [*DOI :* 10.1007/978-3-030-17184-1_15], https://hal.archives-ouvertes.fr/hal-02415758

[28] A. CONDOLUCI, B. ACCATTOLI, C. S. COEN.*Sharing Equality is Linear*, in "PPDP 2019 - 21st International Symposium on Principles and Practice of Programming Languages", Porto, Portugal, October 2019 [*DOI :* 10.1145/3354166.3354174], https://hal.archives-ouvertes.fr/hal-02415769

Books or Proceedings Editing

[29] B. ACCATTOLI, C. OLARTE (editors). *Preface*, The proceedings of LSFA 2018, the 13th Workshop on Logical and Semantic Frameworks with Applications (LSFA'18), Elsevier, August 2019, vol. 344, p. 1-2 [*DOI :* 10.1016/J.ENTCS.2019.07.001], https://hal.archives-ouvertes.fr/hal-02415802

Other Publications

[30] S. MARIN, M. MORALES, L. STRASSBURGER.*A fully labelled proof system for intuitionistic modal logics*, September 2019, working paper or preprint, https://hal.inria.fr/hal-02390454

[31] A. A. TUBELLA, L. STRASSBURGER.*Introduction to Deep Inference*, August 2019, Lecture, https://hal.inria.fr/hal-02390267

References in notes

[32] S. ABRAMSKY.*Computational Interpretations of Linear Logic*, in "Theoretical Computer Science", 1993, vol. 111, p. 3–57

[33] B. ACCATTOLI, U. DAL LAGO.*Beta reduction is invariant, indeed*, in "Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), CSL-LICS '14, Vienna, Austria, July 14 - 18, 2014", 2014, p. 8:1–8:10, http://doi.acm.org/10.1145/2603088.2603105

[34] J.-M. ANDREOLI.*Logic Programming with Focusing Proofs in Linear Logic*, in "Journal of Logic and Computation", 1992, vol. 2, n° 3, p. 297–347

[35] D. BAELDE, D. MILLER, Z. SNOW.*Focused Inductive Theorem Proving*, in "Fifth International Joint Conference on Automated Reasoning (IJCAR 2010)", J. GIESL, R. HÄHNLE (editors), LNCS, 2010, n^{o} 6173, p. 278–292 [*DOI :* 10.1007/978-3-642-14203-1], http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/ijcar10.pdf

[36] G. E. BLELLOCH, J. GREINER.*Parallelism in Sequential Functional Languages*, in "Proceedings of the seventh international conference on Functional programming languages and computer architecture, FPCA 1995, La Jolla, California, USA, June 25-28, 1995", 1995, p. 226–237, http://doi.acm.org/10.1145/224164.224210

[37] K. CHAUDHURI.*The Focused Inverse Method for Linear Logic*, Carnegie Mellon University, December 2006, Technical report CMU-CS-06-162, http://reports-archive.adm.cs.cmu.edu/anon/2006/CMU-CS-06-162.pdf

[38] K. CHAUDHURI, N. GUENOT, L. STRASSBURGER.*The Focused Calculus of Structures*, in "Computer Science Logic: 20th Annual Conference of the EACSL", Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, September 2011, p. 159–173 [*DOI :* 10.4230/LIPIcs.CSL.2011.159], http://drops.dagstuhl.de/opus/volltexte/2011/3229/pdf/16.pdf

[39] K. CHAUDHURI, S. HETZL, D. MILLER.*A Multi-Focused Proof System Isomorphic to Expansion Proofs*, in "Journal of Logic and Computation", June 2014 [*DOI :* 10.1093/LOGCOM/EXU030], http://hal.inria.fr/hal-00937056

[40] A. GACEK, D. MILLER, G. NADATHUR.*Combining generic judgments with recursive definitions*, in "23th Symp. on Logic in Computer Science", F. PFENNING (editor), IEEE Computer Society Press, 2008, p. 33–44, http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/lics08a.pdf

[41] A. GACEK, D. MILLER, G. NADATHUR.*Nominal abstraction*, in "Information and Computation", 2011, vol. 209, n^{o} 1, p. 48–73, http://arxiv.org/abs/0908.1390

[42] J.-Y. GIRARD.*Linear Logic*, in "Theoretical Computer Science", 1987, vol. 50, p. 1–102

[43] S. GRAHAM-LENGRAND, R. DYCKHOFF, J. MCKINNA.*A Focused Sequent Calculus Framework for Proof Search in Pure Type Systems*, in "Logical Methods in Computer Science", 2011, vol. 7, n^{o} 1, http://www.csl.sri.com/users/sgl/Work/Reports/TTSC09.pdf

[44] A. GUGLIELMI.*A System of Interaction and Structure*, in "ACM Trans. on Computational Logic", 2007, vol. 8, n^{o} 1

[45] A. GUGLIELMI, T. GUNDERSEN.*Normalisation Control in Deep Inference Via Atomic Flows*, in "Logical Methods in Computer Science", 2008, vol. 4, n^{o} 1:9, p. 1–36, http://arxiv.org/abs/0709.1205

[46] A. GUGLIELMI, L. STRASSBURGER.*Non-commutativity and MELL in the Calculus of Structures*, in "Computer Science Logic, CSL 2001", L. FRIBOURG (editor), LNCS, Springer-Verlag, 2001, vol. 2142, p. 54–68

[47] D. HUGHES.*Proofs Without Syntax*, in "Annals of Mathematics", 2006, vol. 164, n^{o} 3, p. 1065–1076

[48] F. LAMARCHE, L. STRASSBURGER.*Naming Proofs in Classical Propositional Logic*, in "Typed Lambda Calculi and Applications, TLCA 2005", P. URZYCZYN (editor), LNCS, Springer, 2005, vol. 3461, p. 246–261

[49] C. LIANG, D. MILLER.*Focusing and Polarization in Linear, Intuitionistic, and Classical Logics*, in "Theoretical Computer Science", 2009, vol. 410, n° 46, p. 4747–4768, http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/tcs09.pdf

[50] C. LIANG, D. MILLER.*A Focused Approach to Combining Logics*, in "Annals of Pure and Applied Logic", 2011, vol. 162, n° 9, p. 679–697, http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/lku.pdf

[51] P. MARTIN-LÖF.*Constructive Mathematics and Computer Programming*, in "Sixth International Congress for Logic, Methodology, and Philosophy of Science", Amsterdam, North-Holland, 1982, p. 153–175

[52] R. MCDOWELL, D. MILLER.*Reasoning with Higher-Order Abstract Syntax in a Logical Framework*, in "ACM Trans. on Computational Logic", 2002, vol. 3, n° 1, p. 80–136, http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/mcdowell01.pdf

[53] R. MCDOWELL, D. MILLER.*A Logic for Reasoning with Higher-Order Abstract Syntax*, in "Proceedings, Twelfth Annual IEEE Symposium on Logic in Computer Science", Warsaw, Poland, G. WINSKEL (editor), IEEE Computer Society Press, July 1997, p. 434–445

[54] D. MILLER.*Forum: A Multiple-Conclusion Specification Logic*, in "Theoretical Computer Science", September 1996, vol. 165, n° 1, p. 201–232

[55] D. MILLER, G. NADATHUR, F. PFENNING, A. SCEDROV.*Uniform Proofs as a Foundation for Logic Programming*, in "Annals of Pure and Applied Logic", 1991, vol. 51, p. 125–157

[56] D. MILLER, A. TIU.*A Proof Theory for Generic Judgments: An extended abstract*, in "Proc. 18th IEEE Symposium on Logic in Computer Science (LICS 2003)", IEEE, June 2003, p. 118–127, http://www.lix.polytechnique.fr/Labo/Dale.Miller/papers/lics03.pdf

[57] F. PFENNING, C. SCHÜRMANN.*System Description: Twelf — A Meta-Logical Framework for Deductive Systems*, in "16th Conference on Automated Deduction", Trento, H. GANZINGER (editor), LNAI, Springer, 1999, n° 1632, p. 202–206

[58] B. PIENTKA, J. DUNFIELD.*Beluga: A Framework for Programming and Reasoning with Deductive Systems (System Description)*, in "Fifth International Joint Conference on Automated Reasoning", J. GIESL, R. HÄHNLE (editors), LNCS, 2010, n° 6173, p. 15–21

[59] E. P. ROBINSON.*Proof Nets for Classical Logic*, in "Journal of Logic and Computation", 2003, vol. 13, p. 777–797

[60] THE COQ DEVELOPMENT TEAM.*The Coq Proof Assistant Version 8.3 Reference Manual*, Inria, October 2010

[61] A. TIU, D. MILLER.*Proof Search Specifications of Bisimulation and Modal Logics for the π-calculus*, in "ACM Trans. on Computational Logic", 2010, vol. 11, n° 2, http://arxiv.org/abs/0805.2785

Project-Team PETRUS

PErsonal & TRUSted cloud

IN PARTNERSHIP WITH:
Université Versailles Saint-Quentin

RESEARCH CENTER
Saclay - Île-de-France

THEME
Data and Knowledge Representation and Processing

Table of contents

Project-Team PETRUS

Creation of the Team: 2016 December 01, updated into Project-Team: 2017 July 01

Keywords:

Computer Science and Digital Science:

A1.1.8. - Security of architectures
A1.4. - Ubiquitous Systems
A3.1.2. - Data management, quering and storage
A3.1.3. - Distributed data
A3.1.5. - Control access, privacy
A3.1.6. - Query optimization
A3.1.8. - Big data (production, storage, transfer)
A3.1.9. - Database
A4.3. - Cryptography
A4.5. - Formal methods for security
A4.7. - Access control
A4.8. - Privacy-enhancing technologies

Other Research Topics and Application Domains:

B2.5.3. - Assistance for elderly
B6.4. - Internet of things
B6.6. - Embedded systems
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Nicolas Anciaux [Team leader, Inria, Senior Researcher, HDR]
Luc Bouganim [Inria, Senior Researcher, HDR]

Faculty Members

Guillaume Scerri [Univ de Versailles Saint-Quentin-en-Yvelines, Associate Professor]
Philippe Pucheral [Univ de Versailles Saint-Quentin-en-Yvelines, Professor, HDR]
Iulian Sandu Popa [Univ de Versailles Saint-Quentin-en-Yvelines, Associate Professor]

External Collaborator

Benjamin Nguyen [INSA CVL, Professor, HDR]

Technical Staff

Aydogan Ersoz [Inria, Engineer, until Aug 2019]
Laurent Schneider [Inria, Engineer]

PhD Students

Robin Carpentier [Univ de Versailles Saint-Quentin-en-Yvelines, PhD Student]
Riad Ladjel [Inria until Sept 2019, UVSQ since Sept 2019, PhD Student]
Julien Loudet [CozyCloud, PhD Student, until Sep 2019, granted by CIFRE]
Axel Michel [INSA CVL, PhD Student, until Apr 2019]
Dimitrios Tsolovos [Inria, PhD Student]

Post-Doctoral Fellow

Julien Loudet [Inria, Post-Doctoral Fellow, from Nov 2019]

Administrative Assistants

Adeline Lochet [Inria, Administrative Assistant, until May 2019]

Emmanuelle Perrot [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

We are witnessing an exponential accumulation of personal data on central servers: data automatically gathered by administrations and companies but also data produced by individuals themselves (e.g., photos, agendas, data produced by smart appliances and quantified-self devices) and deliberately stored in the cloud for convenience. The net effect is, on the one hand, an unprecedented threat on data privacy due to abusive usage and attacks and, on the other hand, difficulties in providing powerful user-centric services (e.g. personal big data) which require crossing data stored today in isolated silos. The Personal Cloud paradigm holds the promise of a Privacy-by-Design storage and computing platform, where each individual can gather her complete digital environment in one place and share it with applications and users, while preserving her control. However, this paradigm leaves the privacy and security issues in user's hands, which leads to a paradox if we consider the weaknesses of individuals' autonomy in terms of computer security, ability and willingness to administer sharing policies. The challenge is however paramount in a society where emerging economic models are all based - directly or indirectly - on exploiting personal data.

While many research works tackle the organization of the user's workspace, the semantic unification of personal information, the personal data analytics problems, the objective of the PETRUS project-team is to tackle the privacy and security challenges from an architectural point of view. More precisely, our objective is to help providing a technical solution to the personal cloud paradox. More precisely, our goals are (i) to propose new architectures (encompassing both software and hardware aspects) for secure personal cloud data management and formally prove important bricks of the architecture, (ii) propose new data administration models reaching the main requirements of a personal cloud (decentralized access and usage control models, data sharing, data collection and retention models, etc.) and study the enforcement of the resulting privacy policies based on secure hardware and formally proven architectural components, (iii) propose new secure distributed database indexing models, privacy preserving query processing strategies and data anonymization techniques for the personal cloud.

3. Research Program

3.1. Research Program

To tackle the challenge introduced above, we identify three main lines of research:

- (Axis 1) Personal cloud server architectures. Based on the intuition that user control, security and privacy are key properties in the definition of trusted personal cloud solutions, our objective is to propose new architectures (encompassing both software and hardware aspects) for secure personal cloud data management and formally prove important bricks of the architecture. We also focus in this axis on administration models and their enforcement in relation to the architecture of the system, so that the exclusive control of a non expert individual can be ensured.
- (Axis 2) Global query evaluation. The goal of this line of research is to provide capabilities for crossing data belonging to multiple individuals (e.g., performing statistical queries over personal data, computing queries on social graphs or organizing participatory data collection) in a fully decentralized setting while providing strong and personalized privacy guarantees. This means proposing new secure distributed database indexing models and query processing strategies. In

addition, we concentrate on locally ensuring to each participant the good behaviour of the processing, such that no collective results can be produced if privacy conditions are not respected by other participants.

- (Axis 3) Economic, legal and societal issues. This research axis is more transverse and entails multidisciplinary research, addressing the links between economic, legal, societal and technological aspects. We will follow here a multi-disciplinary approach based on a 3-step methodology: i) identifying important common issues related to privacy and to the exploitation of personal data; ii) characterizing their dimensions in all relevant disciplines and jointly study their entanglement; iii) validating the proposed analysis, models and trade-offs thanks to in vivo experiments.

These contributions will also rely on tools (algorithms, protocols, proofs, etc.) from other communities, namely security (cryptography, secure multiparty computations, formal methods, differential privacy, etc.) and distributed systems (distributed hash tables, gossip protocols, etc.). Beyond the research actions, we structure our software activity around a single common platform (rather than isolated demonstrators), integrating our main research contributions, called PlugDB. This platform is the cornerstone to help validating our research results through accurate performance measurements on a real platform, a common practice in the DB community, and target the best conferences. It is also a strong vector to federate the team, simplify the bootstrapping of new PhD or master students, conduct multi-disciplinary research and open the way to industrial collaborations and technological transfers.

4. Application Domains

4.1. Personal cloud, home care, IoT, sensing, surveys

As stated in the software section, the Petrus research strategy aims at materializing its scientific contributions in an advanced hardware/software platform with the expectation to produce a real societal impact. Hence, our software activity is structured around a common Secure Personal Cloud platform rather than several isolated demonstrators. This platform will serve as the foundation to develop a few emblematic applications. Several privacy-preserving applications can actually be targeted by a Personal Cloud platform, like: (i) smart disclosure applications allowing the individual to recover her personal data from external sources (e.g., bank, online shopping activity, insurance, etc.), integrate them and cross them to perform personal big data tasks (e.g., to improve her budget management) ; (ii) management of personal medical records for care coordination and well-being improvement; (iii) privacy-aware data management for the IoT (e.g., in sensors, quantified-self devices, smart meters); (iv) community-based sensing and community data sharing; (v) privacy-preserving studies (e.g., cohorts, public surveys, privacy-preserving data publishing). Such applications overlap with all the research axes described above but each of them also presents its own specificities. For instance, the smart disclosure applications will focus primarily on sharing models and enforcement, the IoT applications require to look with priority at the embedded data management and sustainability issues, while community-based sensing and privacy-preserving studies demand to study secure and efficient global query processing. Among these applications domains, one is already receiving a particular attention from our team. Indeed, we gained a strong expertise in the management and protection of healthcare data through our past DMSP (Dossier Medico-Social Partagé) experiment in the field. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. We are currently deploying 10000 boxes equipped with PlugDB in the context of the DomYcile project. In this context, we are currently setting up an Inria Innovation Lab with the Hippocad company to industrialize this platform and deploy it at large scale (see Section the bilateral contract OwnCare II-Lab).

5. New Software and Platforms

5.1. PLUG-DB ENGINE

KEYWORDS: Databases - Personal information - Privacy - Hardware and Software Platform

FUNCTIONAL DESCRIPTION: en PlugDB is a complete platform dedicated to a secure and ubiquitous management of personal data. It aims at providing an alternative to a systematic centralization of personal data. The PlugDB engine is a personal database server capable of storing data (tuples and documents) in tables and BLOBs, indexing them, querying them in SQL, sharing them through assertional access control policies and enforcing transactional properties (atomicity, integrity, durability).

The PlugDB engine is embedded in a tamper-resistant hardware device combining the security of smartcard with the storage capacity of NAND Flash. The personal database is hosted encrypted in NAND Flash and the PlugDB engine code runs in the microcontroller. Complementary modules allow to pre-compile SQL queries for the applications, communicate with the DBMS from a remote Java program, synchronize local data with remote servers (typically used for recovering the database in the case of a broken or lost devices) and participate in distributed computation (e.g., global queries). PlugDB runs both on secure devices provided by Gemalto and on specific secure devices designed by PETRUS and assembled by electronic SMEs. Mastering the hardware platform opens up new research and experiment opportunities (e.g., support for wireless communication, secure authentication, sensing capabilities, battery powered ...). PlugDB engine has been registered first at APP (Agence de Protection des Programmes) in 2009 - a new version being registered every two years - and the hardware datasheets in 2015.

PlugDB has been experimented in the field, notably in the healthcare domain. We also recently set up an educational platform on top of PlugDB, named SIPD (Système d'Information Privacy-by-Design) and used at ENSIIE, INSA CVL and UVSQ through the Versailles Sciences Lab fablab, to raise students awareness of privacy protection problems and embedded programming. As a conclusion, PlugDB combines several research contributions from the team, at the crossroads of flash data management, embedded data processing and secure distributed computations. It then strongly federates all members of our team (permanent members, PhD students and engineers). It is also a vector of visibility, technological transfer and dissemination and gives us the opportunity to collaborate with researchers from other disciplines around a concrete privacy-enhancing platform.

PlugDB is now being industrialized in the context of the OwnCare Inria Innovation Lab (II-Lab). In OwnCare, PlugDB acts as a secure personal cloud to manage medical/social data for people receiving care at home. It should be deployed over 10.000 patient in the Yvelines district. The industrialization process covers the development of a complete testing environment, the writing of a detailed documentation and the development of additional features (e.g., embedded ODBC driver, TPM support, flexible access control model and embedded code upgrade notably). It has also required the design of a new hardware platform equipped with a battery power supply, introducing new energy consumption issues for the embedded software.

- Participants: Aydogan Ersoz, Laurent Schneider, Luc Bouganim, Nicolas Anciaux and Philippe Pucheral
- Contact: Nicolas Anciaux
- URL: https://project.inria.fr/plugdb/

6. New Results

6.1. The Security Properties of a PDMS (Axis 1)

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Philippe Pucheral, Iulian Sandu Popa, Guillaume Scerri.

Different Personal Data Management Systems (PDMS) solutions are emerging in both academia and industry. In terms of functionality and security properties, PDMS solutions differ significantly from traditional Data Base Management Systems (DBMS). In a journal article published in Information Systems this year [3], we take stock of the functionality and security of PDMS solutions, propose five very specific security properties to be achieved and provide a preliminary architecture to meet them based on secure hardware [3]. We also presented as a tutorial at VLDB'19 [4] and a keynote at APVP'19 a review of the literature on database and security on data management issues for secure hardware and new research directions for privacy preserving management of personal data.

6.2. SEP2P and DISPERS (Axis 2)

Participants: Luc Bouganim [correspondent], Julien Loudet, Iulian Sandu Popa.

Personal Data Management Systems (PDMS) arrive at a rapid pace allowing us to integrate all our personal data in a single place and use it for our benefit and for the benefit of the community. This leads to a significant paradigm shift since personal data become massively distributed and opens an important question: how can users/applications execute queries and computations over this massively distributed data in a secure and efficient way, relying exclusively on peer-to-peer (P2P) interactions despite covert adversaries which could be executing the query? We first proposed a Secure and Efficient Peer-to-Peer protocol (SEP2P) to randomly select the nodes that will execute the query. This protocol leverages properties of distributed hash tables (DHT) to select nodes in a way that is, at the same time, secure, random and efficient. The security and randomness stem from the fact that we know, with a very high probability, that at least one honest node contributed to the creation and attestation of this list of nodes; while the efficiency stems from the fact that very few nodes are involved in this process. Building on top of SEP2P, we designed DISPERS, a protocol that applies three design rules: (D1) imposed randomness, enforced by SEP2P, (D2) knowledge dispersion, and (D3) task compartmentalization: Each user provides profile information to indexing nodes, chosen randomly thanks to the DHT (D1). Shamir secret-sharing techniques are used to avoid that any indexing node has a full knowledge of indexed nodes (D2). Then, for each query, a set of random nodes is selected (SEP2P) to coordinate the research for query targets using the indexing nodes. Each of these random nodes learns a part of the query targets IP address but does not know the query (D2, D3). Another set of random nodes is chosen to compute of the final answer based on partial local results from targets. These nodes learn part of the results but do not know the targets, thanks to proxies, nor the meaning of these results (D2, D3). These results are the core of Julien Loudet's thesis [1]. SEP2P was published at EDBT'19 [9] while a demonstration of DISPERS was published at VLDB'19 [8].Both works were also exposed/demonstrated at BDA'19 [13] [12] and APVP'19 [14] for the French research community in databases and security and privacy.

6.3. Manifest-based Framework for Secure Decentralized Queries (Axis 2)

Participants: Riad Ladjel [correspondent], Nicolas Anciaux, Philippe Pucheral, Guillaume Scerri.

The PDMS context calls for a new decentralized way of handling processing. The challenge is to allow generic treatment of large populations of PDMS, with a double objective: to preserve the mutual trust of individuals in their PDMS, and to guarantee an honest result (calculated on the right data, with the right code). To achieve this goal, our approach introduces a computational 'manifest', stipulating its execution plan and the privacy clauses (e.g., collection rules) to be guaranteed at runtime, based on trusted hardware (e.g., Intel SGX processor). Our contributions consist of (1) a protocol for randomly assigning compute tasks to participants to prevent targeted attacks, (2) a mechanism guaranteeing global compute integrity through local-only checks (without centralized trusted third party) and (3) database countermeasures limiting the impact of hidden channel attacks from corrupted participants. These contributions resulted in articles in TrustCom'19 [7] and ISD'19 [6]. Our approach guarantees confidentiality and processing integrity, it is generic and scalable, and goes far beyond existing approaches (e.g., secure multiparty computing or differential privacy).

6.4. Mobile Participatory Sensing with Strong Privacy Guarantees (Axis 2)

Participant: Iulian Sandu Popa [correspondent].

Mobile participatory sensing (MPS) could benefit many application domains. A major domain is smart transportation, with applications such as vehicular traffic monitoring, vehicle routing, or driving behavior analysis. However, MPS's success depends on finding a solution for querying large numbers of smart phones or vehicular systems, which protects user location privacy and works in real-time. This work proposes PAMPAS, a privacy-aware mobile distributed system for efficient data aggregation in MPS. In PAMPAS, mobile devices enhanced with secure hardware, called secure probes (SPs), perform distributed query processing, while preventing users from accessing other users' data. A supporting server infrastructure (SSI) coordinates the inter-SP communication and the computation tasks executed on SPs. PAMPAS ensures that SSI cannot link the location reported by SPs to the user identities even if SSI has additional background information. Moreover, an enhanced version of the protocol, named PAMPAS$^+$, makes the system robust even against advanced hardware attacks on the SPs. Hence, the risk of user location privacy leakage remains very low even for an attacker controlling the SSI and a few corrupted SPs. Our experimental results demonstrate that these protocols work efficiently on resource constrained SPs being able to collect the data, aggregate them, and share statistics or derive models in real-time. This work has been accomplished in collaboration with NJIT and DePaul University and has been recently accepted as a journal paper (an 'Online first' version is available at https://link.springer.com/article/10.1007/s10707-019-00389-4).

6.5. Empowerment and Big Data on Personal Data: from Portability to Agency (Axis 3)

Participants: Nicolas Anciaux [correspondent], Riad Ladjel, Philippe Pucheral, Guillaume Scerri.

The current highly centralised model of personal data management is based on established business practices that have led to widespread adoption, in contrast to user-centric and privacy-oriented systems such as PDMS, which therefore need to be studied in terms of technical, economic and legal feasibility and adoptability with researchers from other disciplines. In the context of the DATAIA GDP-ERE project, we are analyzing the technical and legal conditions under which individuals can exercise their right to data portability. Over the period, we have jointly studied a new notion that characterizes the true power of the individual over his or her personal data: agency. In particular, we have shown how the notion of agency, which comes from the social sciences, can be transposed and used to our context to measure the empowerment of individuals in Big Data applications. This study led to two joint publications with law researchers over the period, in particular in the journal Dalloz IP/IT [5], as well as several international panels (see in Section Popularization, e.g., panel at BDVA forum organized in Helsinky with the European Commission, at the Annual Forum of Trans Europ Expert, etc.)

6.6. OwnCare Inria Inovation Lab

Participants: Philippe Pucheral [correspondent], Nicolas Anciaux, Luc Bouganim, Laurent Schneider.

The OwnCare IILab was created in January 2018 (see section: Bilateral Contracts with Industry) and involves the Hippocad SME and the PETRUS team around the management of medical-social data at patient's home. The objective is to build a fully decentralized and highly secure personal medical-social folder based on PlugDB, and deploy it at large scale. Besides this industrial objective, the goal is also to leverage and validate the PETRUS research contributions related to secured Personal Cloud architectures. Before the creation of the OwnCare IILab initiative, PlugDB was an advanced research prototype. It is now evolving towards a transferable product. To reach this state, a considerable effort has been made in terms of development, testing platform, validation procedures and documentation. PlugDB engine is regularly registered at APP (Agence de Protection des Programmes), for both the PlugDB hardware datasheets and the code of the PlugDB-engine. The next PlugDB code registration will cover all functionalities added since the beginning of the IILab, notably: dynamic upgrade of the embedded code, TPM-based secure boot, ad-hoc embedded stored procedures, RBAC-style access control model, aggregate computation, SSL certificate management, event/error logging mechanism. Some of these developments are highly challenging considering the embedded context and the energy consumption constraints we have to face (the current device hosting PlugDB is based on two microcontrollers – MCU – powered by small batteries). Typically, we had to implement the first coupling between a TPM and a STM MCU, a lightweight version of SSL that accommodates MCU resources and energy-saving synchronization protocols between 2 MCU.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. OwnCare II-Lab (Jul 2017 - Dec 2020)

Partners: PETRUS (Inria-UVSQ), Hippocad (SME)

End 2016, the Yvelines district lauched a public call for tender to deploy an industrial solution aiming at covering the whole distinct (10.000 patients). The Hippocad company, in partnership with Inria, won this call for tender with a solution called DomYcile in May 2017 and the project was launched in July 2017. DomYcile is based on a home box combining the PlugDB hardware/software technology developed by the Petus team and a communication layer based on SigFox. Hippocad and Petrus then decided to launch a joint II-Lab (Inria Innovation Lab) named OwnCare. The objective is threefold: (1) build an industrial solution based on PlugDB and deploy it in the Yvelines district in the short-term, (2) use this Yvelines testbed to improve the solution and try to deploy it at the national/international level in the medium-term and (3) design flexible/secure/mobile personal medical folder solutions targeting individual uses rather than professional uses in the long-term. The DomYcile project with the Yvelines district has started in July 2017 and the II-Lab was officially created in January 2018.

7.2. Bilateral Grants with Industry

7.2.1. Cozy Cloud CIFRE - Loudet contract (Apr 2016 - Apr 2019)

Partners: Cozy Cloud, PETRUS

In relation with the bilateral contract mentioned above, a second CIFRE PhD thesis has been started by Julien Loudet. The objective is to allow for a secure execution of distributed queries on a set of personal clouds associated to users, depending on social links, user's localization or user's profile. The general idea is to build secure indexes, distributed on the users' personal clouds and to devise a secure execution protocol revealing solely the query result to the querier. Such highly distributed secure queries potentially enable new (social) applications fed by user's personal data which could be developed on the Cozy-PlugDB platform.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR PerSoCloud (Jan 2017 - Dec 2020)

Partners: Orange Labs (coordinator), PETRUS (Inria-UVSQ), Cozy Cloud, U. of Versailles.

The objective of PerSoCloud is to design, implement and validate a full-fledged Privacy-by-Design Personal Cloud Sharing Platform. One of the major difficulties linked to the concept of personal cloud lies in organizing and enforcing the security of the data sharing while the data is no longer under the control of a central server. We identify three dimensions to this problem. Devices-sharing: assuming that the primary copy of user U1's personal data is hosted in a secure place, how to share and synchronize it with U1's multiple (mobile) devices without compromising security? Peers-sharing: how user U1 could exchange a subset of his-her data with an identified user U2 while providing to U1 tangible guarantees about the usage made by U2 of this data? Community-sharing: how user U1 could exchange a subset of his-her data with a large community of users and contribute to personal big data analytics while providing to U1 tangible guarantees about the preservation of his-her anonymity? In addition to tackling these three scientific and technical issues, a legal analysis will guarantee compliance of this platform with the security and privacy French and UE regulation, which firmly promotes the Privacy by Design principle, including the current reforms of personal data regulation.

8.1.2. GDP-ERE, DATA-IA project (Sept. 2018 - Jan. 2022)

Partners: DANTE (U. of Versailles), PETRUS (Inria-UVSQ).

The role of individuals and the control of their data is a central issue in the new European regulation (GDPR) enforced on 25th May 2018. Data portability is a new right provided under those regulations. It allows citizens to retrieve their personal data from the companies and governmental agencies that collected them, in an interoperable digital format. The goals are to enable the individual to get out of a captive ecosystem, and to favor the development of innovative personal data services beyond the existing monopolistic positions. The consequence of this new right is the design and deployment of technical platforms, commonly known as Personal Cloud. But personal cloud architectures are very diverse, ranging from cloud based solutions where millions of personal cloud are managed centrally, to self-hosting solutions. These diversity is not neutral both in terms of security and from the point of view of the chain of liabilities. The GDP-ERE project tends to study those issues in an interdisciplinary approach by the involvement of jurists and computers scientists. The two main objectives are (i) to analyze the effects of the personal cloud architectures on legal liabilities, enlightened by the analysis of the rules provided under the GDPR and (ii) to propose legal and technological evolutions to highlight the share of liability between each relevant party and create adapted tools to endorse those liabilities. http://dataia.eu/actualites/linstitut-dataia-vous-presente-le-projet-gdp-ere-rgpd-et-cloud-personnel-de-lempowerment

8.1.3. Postdoc DIM RFSI, Ile-de-France Region (2019 - 2020)

Partners: Inria (PETRUS).

This project is a continuation of Julien Loudet's Phd thesis. Julien finalized a CIFRE thesis defended in October 2019. This thesis is the result of a solid collaboration (another CIFRE thesis was defended in 2018) between the PETRUS team and the startup Cozy Cloud, which is also working on the personal cloud issue. The project finances 8 months of postdoc for Julien. The objective is to enforce the collaboration with Cozy Cloud by allowing the postdoc (i) to submit an extended journal paper on his last results (DISPERS protocol), (ii) to realize a detailed specification of the distributed protocols developed during his PhD for their implementation in the Cozy Cloud platform and (iii) to collaborate with a future PhD candidate of a new thesis in collaboration with Cozy Cloud exploring decentralized automatic learning techniques in the personal cloud context.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

- Luc Bouganim: Co-organizer of Ecole thématique BDA Masses de Données Distribuées, Aussois, June 2020
- Iulian Sandu Popa: Colloque National Capteurs et Sciences Participatives (CASPA), Paris, 1-4 avril 2019

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

- Nicolas Anciaux: VLDB'19, SIGMOD'19
- Luc Bouganim: EDBT'2020
- Philippe Pucheral: DATA'19, MOBILITY'19, MEDES'19
- Iulian Sandu Popa: ICDE'20, SSDBM'19, IEEE MobileCloud'19, DATA'19, BDA'19 demo
- Guillaume Scerri: BDA'19

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Nicolas Anciaux: Associate Editor of the VLDB Journal

9.1.3.2. Reviewer - Reviewing Activities

- Iulian Sandu Popa: Journal of Internet Services and Applications
- Guillaume Scerri: ACM Transactions on Privacy and Security

9.1.4. Research Administration

- Philippe Pucheral: Member of the HDR committee of the STV doctoral school (UVSQ) since 2014
- Philippe Pucheral: Member of the steering committee of the ED STIC doctoral school of University Paris-Saclay, 'Data, Knowledge and Interactions' committee (about 250 PhD students) since 2014
- Philippe Pucheral: Member of the bureau of the DAVID lab board since 2016
- Nicolas Anciaux: Member of the Council of the Doctoral College of the University Paris-Saclay
- Nicolas Anciaux: Member of the "Bureau du comité des projets" at Inria saclay
- Nicolas Anciaux: Member of the DATAIA Convergence Institute Program Committee since 2018.
- Nicolas Anciaux: Correspondent for the Doctoral school ED STIC of University Paris-Saclay at Inria Saclay
- Nicolas Anciaux: Responsible for the 'Mission Jeunes Chercheurs' (MJC) at Inria Saclay
- Nicolas Anciaux: Responsible for the 'Formation par la Recherche' (FPR) at Inria Saclay
- Nicolas Anciaux: Member of the bureau of the DAVID lab board
- Luc Bouganim: Member of the Scientific Commission (CS) of Inria Saclay-IDF (Cordi-S, Post-Doc, Delegation)
- Luc Bouganim: Member of the Commission for Technological Development (CDT) of Inria Saclay-IDF
- Luc Bouganim: Member of the admission committee for DR2 at Inria
- Luc Bouganim: Member of the recruitment committee for CRCN at Inria
- Luc Bouganim: Member of the recruitment committee of associate professor at ITU Copenhagen
- Iulian Sandu Popa: Member of the Commission for Technological Development (CDT) of Inria Saclay-IDF
- Iulian Sandu Popa: Member of the recruitment committee for an MCF (Maître de Conférences) position at CNAM
- Iulian Sandu Popa: PhD thesis referent for the ED STIC doctoral school of University Paris-Saclay
- Guillaume Scerri: PhD thesis referent for the ED STIC doctoral school of University Paris-Saclay

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Licence : Iulian Sandu Popa, Bases de données (niveau L3), 96, UVSQ, France. Guillaume Scerri, Initiation aux bases de données (niveau L2), 63, UVSQ, France. Guillaume Scerri, Fondements de l'informatique (niveau L1), 36, UVSQ, France. Guillaume Scerri, Théorie des Langages (niveau L2), 45, UVSQ, France.
- Master : Iulian Sandu Popa, Bases de données relationnelles (niveau M1), Gestion des données spatiotemporelles (niveau M2), Sécurité des bases de données (niveau M2), 96, UVSQ, France. Philippe Pucheral, responsible of the DataScale master, courses in M1 and M2 in databases and in security, introductory courses for jurists,UVSQ, France. Guillaume Scerri, Bases de données relationnelles (niveau M1), 36, UVSQ, France. Guillaume Scerri, Sécurité et bases de données pour juristes, 4.5, UVSQ, France. Guillaume Scerri, Sécurité, 18, UVSQ, France.

- Engineers school : Nicolas Anciaux, courses on Databases (module IN206, niveau M1), 21, and Advanced databases (module ASI13, niveau M2), 24, at ENSTA ParisTech. Nicolas Anciaux, Systèmes d'Information "privacy by design" (niveau M1), 30, at ENSIIE Evry, France. Luc Bouganim, Systèmes d'Information "privacy by design" (niveau M1), 42, ENSIIE Evry, France. Luc Bouganim, Bases de données relationnelles (niveau M1), 32, ENSTA, France.

9.2.2. Supervision

- PhD : Julien Loudet, Distributed and Privacy-Preserving Personal Queries on Personal Clouds, UVSQ, October 2019, Luc Bouganim and Iulian Sandu Popa
- PhD: Axel Michel, Secure Distributed Computations, INSA CVL, April 8, 2019, Benjamin Nguyen and Philippe Pucheral
- PhD in progress: Riad Ladjel, Secure Distributed Computation for the Personal Cloud, October 2016, Nicolas Anciaux, Philippe Pucheral and Guillaume Scerri
- PhD in progress: Dimitris Tsoulovos, Privacy-by-design Middleware for Urban-scale Mobile Crowdsensing, April 2017, Nicolas Anciaux and Valérie Issarny (Inria Mimove)
- PhD in progress: Robin Carpentier, Secure and efficient data processing in trusted execution environments for the personal cloud, October 2018, Nicolas Anciaux, Iulian Sandu Popa and Guillaume Scerri

9.2.3. Juries

- Nicolas Anciaux : Jury Member of the PhD of Rémy Delanaux (University Lyon I)
- Philippe Pucheral : President of the PhD jury of Karima RAFES (Paris-Saclay University, 25/01/2019)
- Philippe Pucheral : President of the PhD jury of Duc CAO (Paris-Saclay University, 26/09/2019)

9.3. Popularization

- Nicolas Anciaux: "Sécurité des données personnelles, démonstration de la box 'DomYcile' développée par PETRUS et Hippocad pour le département des Yvelines, Journées du Patrimoine 2019 à l'Inria Rocquencourt, Sept. 2019.
- Nicolas Anciaux: Demonstration of the 'DomYcile' home-box, Inria stand, European Big Data Value Forum (EBDVF19), organized by BDVA in collaboration of the European Commission, Oct. 2019.
- Round table on "Personal Dataspaces" at European Big Data Value Forum (EBDVF19), organized by BDVA in collaboration of the European Commission, Oct. 2019. The round table is chaired by Marko Turpeinen (Director EIT Digital Finland). Intervention of N. Anciaux about "Empowerment and Big Data on personal data: from data portability to personal agency".
- Nicolas Anciaux: séminaire LINC à la CNIL. "Solutions de cloud personnel : sécurité, vie privée, et perspectives", Nov. 2019.
- Nicolas Anciaux: "PETRUS : Cloud personnel de confiance et gestion de nos données", Café des Sciences d'Inria, site de Rocquencourt, 11 juin 2019
- N. Anciaux: Panel on "European Data Protection Law", 10th Annual Forum of Trans Europ Expert 'Le droit de l'Union européenne vu d'ailleurs', Conseil Suprérieur du Notariat, Paris. April 2019. Round table chaired by J. Sénéchal (Dir. pôle "Droit des contrats" at TEE) and N. Martial-Braz (Dir. pôle "Droit de la propriété intellectuelle at TEE), with B. Amaudric du Chaffaut (Deputy General Counsel at Google France) on GDPR evolution in Europe, F. Nicola (Prof. at Washington College of Law) on the echo of GDPR in California and N. Anciaux (researcher at Inria) on technical-juridical challenges of the GDPR.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] J. LOUDET.*Distributed and Privacy-Preserving Personal Queries on Personal Clouds*, Université de Versailles Saint Quentin en Yvelines (UVSQ), October 2019, https://hal.inria.fr/tel-02376516

[2] A. MICHEL.*Personalizing Privacy Constraints in Generalization-based Anonymization Models*, INSA Centre Val de Loire, April 2019, https://hal.archives-ouvertes.fr/tel-02269565

Articles in International Peer-Reviewed Journal

[3] N. ANCIAUX, P. BONNET, L. BOUGANIM, B. NGUYEN, P. PUCHERAL, I. SANDU-POPA, G. SCERRI.*Personal Data Management Systems: The security and functionality standpoint*, in "Information Systems", 2019, vol. 80, p. 13-35 [*DOI :* 10.1016/J.IS.2018.09.002], https://hal.archives-ouvertes.fr/hal-01898705

[4] N. ANCIAUX, L. BOUGANIM, P. PUCHERAL, I. S. POPA, G. SCERRI.*Personal Database Security and Trusted Execution Environments: A Tutorial at the Crossroads*, in "Proceedings of the VLDB Endowment (PVLDB)", August 2019, https://hal.inria.fr/hal-02269292

[5] N. ANCIAUX, C. ZOLYNSKI.*Empowerment and Big Data on personal data : from portability to agency*, in "Dalloz IP/IT", 2019, forthcoming, https://hal.inria.fr/hal-02349274

International Conferences with Proceedings

[6] R. LADJEL, N. ANCIAUX, P. PUCHERAL, G. SCERRI.*A manifest-based framework for organizing the management of personal data at the edge of the network*, in "ISD 2019 - 28th International Conference on Information Systems Development", Toulon, France, August 2019, https://hal.archives-ouvertes.fr/hal-02269203

[7] R. LADJEL, N. ANCIAUX, P. PUCHERAL, G. SCERRI.*Trustworthy Distributed Computations on Personal Data Using Trusted Execution Environments*, in "TrustCom 2019 - The 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications / BigDataSE 2019 - 13th IEEE International Conference on Big Data Science and Engineering", Rotorua, New Zealand, August 2019, https://hal.archives-ouvertes.fr/hal-02269207

[8] J. LOUDET, I. SANDU-POPA, L. BOUGANIM.*DISPERS: Securing Highly Distributed Queries on Personal Data Management Systems*, in "VLDB 2019 - 45th International Conference on Very Large Data Bases", Los Angeles, United States, Proceedings of the VLDB Endowment, August 2019, vol. 12, n^{o} 12, 4, https://hal.inria.fr/hal-02269209

[9] J. LOUDET, I. SANDU-POPA, L. BOUGANIM.*SEP2P: Secure and Efficient P2P Personal Data Processing*, in "EDBT 2019 - 22nd International Conference on Extending Database Technology", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-01949641

Conferences without Proceedings

[10] R. LADJEL, N. ANCIAUX, P. PUCHERAL, G. SCERRI.*Trustworthy Distributed Computations on Personal Data Using Trusted Execution Environments*, in "APVP 2019 - Atelier sur la Protection de la Vie Privée", Cap Hornu, France, July 2019, https://hal.archives-ouvertes.fr/hal-02269200

[11] R. LADJEL, N. ANCIAUX, P. PUCHERAL, G. SCERRI.*Trustworthy Distributed Computations on Personal Data Using Trusted Execution Environments*, in "BDA 2019 - 35ème Conférence sur la Gestion de Données - Principes, Technologies et Applications", Lyon, France, October 2019, https://hal.archives-ouvertes.fr/hal-02269211

[12] J. LOUDET, I. SANDU-POPA, L. BOUGANIM.*DISPERS: Securing Highly Distributed Queries on Personal Data Management Systems*, in "BDA 2019 - 35ème Conférence sur la Gestion de Données - Principes, Technologies et Applications", Lyon, France, October 2019, https://hal.inria.fr/hal-02269201

[13] J. LOUDET, I. SANDU-POPA, L. BOUGANIM.*SEP2P: Secure and Efficient P2P Personal Data Processing*, in "BDA 2019 - 35ème Conférence sur la Gestion de Données - Principes, Technologies et Applications", Lyon, France, October 2019, https://hal.inria.fr/hal-02269222

[14] J. LOUDET, I. SANDU-POPA, L. BOUGANIM.*SEP2P: Secure and Efficient P2P Personal Data Processing*, in "APVP 2019 - Atelier sur la Protection de la Vie Privée", Cap Hornu, France, July 2019, https://hal.inria.fr/hal-02269216

Project-Team POEMS

Wave propagation: mathematical analysis and simulation

IN COLLABORATION WITH: Propagation des ondes : étude mathématique et simulation (POEMS)

IN PARTNERSHIP WITH:
CNRS

Ecole nationale supérieure des techniques avancées

RESEARCH CENTER
Saclay - Île-de-France

THEME
Numerical schemes and simulations

Table of contents

Project-Team POEMS

Creation of the Project-Team: 2019 November 01

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.1.6. - Fractal Modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.7. - High performance computing
A6.3.1. - Inverse problems
A6.3.4. - Model reduction
A6.5.1. - Solid mechanics
A6.5.2. - Fluid mechanics
A6.5.4. - Waves

Other Research Topics and Application Domains:

B2.6. - Biological and medical imaging
B3.3. - Geosciences
B3.3.1. - Earth and subsoil
B3.4. - Risks
B3.4.1. - Natural risks
B3.4.2. - Industrial risks and waste
B5.3. - Nanotechnology
B5.4. - Microelectronics
B5.5. - Materials

1. Team, Visitors, External Collaborators

Research Scientists

Anne-Sophie Bonnet-Ben Dhia [Team leader, CNRS, Senior Researcher, HDR]
Marc Bonnet [CNRS, Senior Researcher, HDR]
Eliane Bécache [Inria, Researcher, HDR]
Stéphanie Chaillat [CNRS, Researcher, HDR]
Christophe Hazard [CNRS, Researcher, HDR]
Patrick Joly [Inria, Senior Researcher, HDR]
Maryna Kachanovska [Inria, Researcher]

Luiz Maltez Faria [Inria, Starting Research Position]
Jean-François Mercier [CNRS, Senior Researcher, HDR]
Axel Modave [CNRS, Researcher]

Faculty Members

Laurent Bourgeois [École Nationale Supérieure de Techniques Avancées, Professor, HDR]
Patrick Ciarlet [École Nationale Supérieure de Techniques Avancées, Professor, HDR]
Sonia Fliss [École Nationale Supérieure de Techniques Avancées, Professor, HDR]
Laure Giovangigli [École Nationale Supérieure de Techniques Avancées, Associate Professor]
Eric Lunéville [École Nationale Supérieure de Techniques Avancées, Professor]

External Collaborator

Francis Collino [Self-employed]

Technical Staff

Colin Chambeyron [CNRS, Engineer]
Nicolas Kielbasiewicz [CNRS, Engineer]
Christophe Mathulik [École Nationale Supérieure de Techniques Avancées, Engineer, until Aug 2019]

PhD Students

Yassine Abourrig [CEA, PhD Student, until March 2019]
Clément Bénéteau [École Nationale Supérieure de Techniques Avancées, PhD Student]
Antoine Bera [Ministère de l'Education Nationale, PhD Student, until Aug 2019]
Damien Chicaud [École Nationale Supérieure de Techniques Avancées, PhD Student]
Damien Mavaleix-Marchessoux [Naval Group, PhD Student]
Hajer Methenni [CEA, PhD Student]
Sandrine Paolantoni [École Nationale Supérieure de Techniques Avancées, PhD Student]
Emile Parolin [Inria, PhD Student, until Oct 2019]
Mahran Rihani [École Nationale Supérieure de Techniques Avancées, PhD Student]
Yohanes Tjandrawidjaja [CEA, PhD Student, until Dec 2019]

Post-Doctoral Fellows

Maria Kazakova [LMH, Post-Doctoral Fellow, until Aug 2019]
Felix Kpadonou [École Nationale Supérieure de Techniques Avancées, Post-Doctoral Fellow, until Aug 2019]
Florian Monteghetti [Inria, Post-Doctoral Fellow, until Sept 2019]

Administrative Assistants

Corinne Chen [École Nationale Supérieure de Techniques Avancées]
Natalia Alves [Inria, until Jan 2019]
Marie Enée [Inria, from May 2019]

2. Overall Objectives

2.1. Overall Objectives

The propagation of waves is one of the most common physical phenomena in nature. From the human scale (sounds, vibrations, water waves, telecommunications, radar) to the scales of the universe (electromagnetic waves, gravity waves) and of the atoms (spontaneous or stimulated emission, interferences between particles), the emission and the reception of waves are our privileged way to understand the world that surrounds us. The study and the simulation of wave propagation phenomena constitute a very broad and active field of research in various domains of physics and engineering sciences. The variety and the complexity of the underlying problems, their scientific and industrial interest, the existence of a common mathematical structure to these problems from different areas altogether justify a research project in applied mathematics and scientific computing devoted to this topic.

Nowadays, the numerical techniques for solving the basic academic problems are well mastered, and significant progress has been made during the last twenty years for handling problems closer to real applications. But several bottlenecks remain, among which one can mention the high-frequency problems for radar applications, the multiscale problems that arise for instance in nanotechnologies or the multi-physics couplings, like in aeroacoustics. Moreover, in the recent period, new challenges have emerged, related to new discoveries in physics (like negative index metamaterials) or to the fantastic development of information and communication techniques. For example, the growing development of increasingly connected objects (internet of things) and the forthcoming availability of autonomous vehicles depend crucially on electromagnetic waves, raising important issues about radar performance, sensor reliability, component miniaturization and electromagnetic compatibility. Generally, there are a lot of application domains which could benefit from advanced research on waves phenomena. Enhancing ultrasound-based methods for detection and imaging, which are already intensively used in e.g. medicine, could permit real-time health monitoring of aircrafts or nuclear plants. Guarding against seismic risks still requires considerable advances in the simulation of elastic waves in large and complex media. And many other applications motivating our research and our prospects could be added to this far-from-comprehensive list.

3. Research Program

3.1. Expertises

The activity of the team is oriented towards the design, the analysis and the numerical approximation of mathematical models for all types of problems involving wave propagation phenomena, in mechanics, physics and engineering sciences. Let us briefly describe our core business and current expertise, in order to clarify the new challenges that we want to address in the short and long terms.

Typically, our works are based on *boundary value problems* established by physicists to model the propagation of waves in various situations. The basic ingredient is a partial differential equation of the hyperbolic type, whose prototype is the scalar wave equation, or the Helmholtz equation if time-periodic solutions are considered. More generally, we systematically consider both the transient problem, in the time domain, and the time-harmonic problem, in the frequency domain. Let us mention that, even if different waves share a lot of common properties, the transition from the scalar acoustic equation to the vectorial electromagnetism and elastodynamics systems raises a lot of mathematical and numerical difficulties, and requires a specific expertise.

A notable particularity of the problems that we consider is that they are generally set in *unbounded domains*: for instance, for radar applications, it is necessary to simulate the interaction of the electromagnetic waves with the airplane only, without any complex environment perturbing the wave phenomena. This raises an intense research activity, both from a theoretical and a numerical point of view. There exist several approaches which all consist in rewriting the problem (or an approximation of it) in a bounded domain, the new formulation being well-suited for classical mathematical and numerical techniques.

One class of methods consists in applying an appropriate condition on some boundary enclosing the zone of interest. In the frequency domain, one can use a non-local transparent condition, which can be expressed by a convolution with a Green function like in integral equation techniques, or by a modal decomposition when a separation of variables is applicable. But for explicit schemes in the time domain, local radiation conditions at a finite distance are generally preferred (constructed as local approximations at various orders of the exact non-local condition). A second class of methods consists in surrounding the computational domain by so called *Perfectly Matched absorbing Layers* (PML), which are very popular because they are easy to implement. POEMS members have provided several contributions to these two classes of methods for more than twenty-five years. Among them, on can mention the understanding of the instability of PMLs in anisotropic media and in dispersive media, the derivation of transparent boundary conditions in periodic media or the improvement of Fast Multipole techniques for elastodynamic integral equations.

In addition to more classical domains of applied mathematics that we are led to use (variational analysis and functional analysis, interpolation and approximation theory, linear algebra of large systems, etc...), we have acquired a deep expertise in *spectral theory*. Indeed, the analysis of wave phenomena is intimately linked to the study of some associated spectral problems. Acoustic resonance frequencies of a cavity correspond to the eigenvalues of a selfadjoint Laplacian operator, modal solutions in a waveguide correspond to a spectral problem set in the cross section. In these two examples, if the cavity or the cross-section is unbounded, a part of the spectrum is a continuum. Again, POEMS has produced several contributions in this field. In particular, a large number of significant results have been obtained for the existence or non-existence of guided modes in open waveguides and of trapped modes in infinite domains.

To end this far from exhaustive presentation of our main expertise domains, let us mention the *asymptotic techniques* with respect to some small scale appearing in the model: it can be the wavelength compared to the size of the scatterer, or on the contrary, the scale of the scatterer compared to the wavelength, it can be the scale of some microstructure in a composite material or the width of a thin layer or a thin tube. In each case, the objective, in order to avoid the use of costly meshes, is to derive effective simplified models. Our specificity here is that we can combine skills in physics, mathematics and numerics: in particular, we take care of the mathematical properties of the effective model, which are used to ensure the robustness of the numerical method, and also to derive error estimates with respect to the small parameter. There has been a lot of contributions of POEMS to this topic, going from the modeling of electromagnetic coatings to the justification of models for piezoelectric sensors. Let us mention that effective models for small scatterers and thin coatings have been used to improve imaging techniques that we are developing (topological gradient, time reversal or sampling techniques).

3.2. New domains

In order to consider more and more challenging problems (involving non-deterministic, large-scale and more realistic models), we decided recently to enlarge our domain of expertise in three directions.

Firstly, we want to reinforce our activity on *efficient solvers for large-scale wave propagation problems*. Since its inception, POEMS has frequently contributed to the development and the analysis of numerical methods that permit the fast solution of large-scale problems, such as high-order finite element methods, boundary elements methods and domain decomposition methods. Nevertheless, implementing these methods in parallel programming environments and dealing with large-scale benchmarks have generally not been done by the team. We want to continue our activities on these methods and, in a more comprehensive approach, we will incorporate modern algebraic strategies and high-performance computing (HPC) aspects in our methodology. In collaboration with academic or industrial partners, we would like to address industrial-scale benchmarks to assess the performance of our approaches. We believe that taking all these aspects into consideration will allow us to design more efficient wave-specific computational tools for large-scale simulations.

Secondly, up to now, *probabilistic methods* were outside the expertise of POEMS team, restricting us to deterministic approaches for wave propagation problems. We however firmly believe in the importance and usefulness of addressing uncertainty and randomness inherent to many propagation phenomena. Randomness may occur in the description of complex propagation media (for example in the modeling of ultrasound waves in concrete for the simulation of non-destructive testing experiments) or of data uncertainties. To quantify the effect of such uncertainties on the design, behavior, performance or reliability of many systems is then a natural goal in diverse fields of application.

Thirdly and lastly, we wish to develop and strengthen collaborations allowing a *closer interaction between our mathematical, modeling and computing activities and physical experiments*, where the latter may either provide reality checks on existing models or strongly affect the choice of modeling assumptions. Within our typical domain of activities, we can mention three areas for which such considerations are highly relevant. One is musical acoustics, where POEMS has made several well-recognized contributions dealing with the simulation of musical instruments. Another area is inverse problems, whose very purpose is to extract useful information from actual measurements with the help of (propagation) models. This is a core of our partnership with CEA on ultrasonic Non Destructive Testing. A third area is the modelling of effective (acoustic or electromagnetic) metamaterials, where predictions based on homogenized models have to be confirmed by experiments.

4. Highlights of the Year

4.1. Highlights of the Year

- The POEMS EPI has been officially renewed by Inria on the 1st November 2019.
- Two permanent members of the team, S. Chaillat and S. Fliss, have successfully defended their habilitation theses.
- SACHEMS project led by CEA (where POEMS is involved) has been retained in the call of Ile-de-France Region SESAME 2019. Its purpose is to federate the research in the region in the field of SHM (Structural Health Monitoring), which consists of developing intelligent sensors aimed at detecting and characterizing directly defects in a structure (crack, corrosion, etc.). The SACHEMS project is part of a strategy for pooling equipment with the objective of creating an innovation platform for SHM methods.

5. New Software and Platforms

5.1. COFFEE

KEYWORDS: Numerical simulations - Wave propagation - Boundary element method

FUNCTIONAL DESCRIPTION: COFFEE is an adapted fast BEM solver to model acoustic and elastic wave propagation (full implementation in Fortran 90). The 3-D acoustic or elastodynamic equations are solved with the boundary element method accelerated by the multi-level fast multipole method or a hierarchical-matrices based representation of the system matrix. The fundamental solutions for the infinite space are used in this implementation. A boundary element-boundary element coupling strategy is also implemented so multi-region problems (strata inside a valley for example) can be solved. In order to accelerate the convergence of the iterative solver, various analytic or algebraic preconditioners are available. Finally, an anisotropic mesh adaptation strategy is used to further reduce the computational times.

- Contact: Stéphanie Chaillat

5.2. XLiFE++

KEYWORDS: Numerical simulations - Finite element modelling - Boundary element method

FUNCTIONAL DESCRIPTION: XLiFE++ is an FEM-BEM C++ code developed by POEMS laboratory and IRMAR laboratory, that can solve 1D/2D/3D, scalar/vector, transient/stationary/harmonic problems. Description: https://uma.ensta-paristech.fr/soft/XLiFE++/

- Contact: Eric Lunéville

6. New Results

6.1. Complex ordered and disordered media

There is a need of a better understanding of wave phenomena in complex media. From a physical point of view, a *complex medium* is typically a material where the propagation of the waves may be *anisotropic* and *dispersive*. These properties are generally the effect of a microstructure, that can be ordered (in e.g. photonic crystals), or disordered (light in the atmosphere, seismic waves). From a mathematical point of view, one can take into account exactly this microstructure or, at sufficiently low frequency, use effective models justified by the *homogenization theory*.

6.1.1. Enriched homogenized model in presence of boundaries for time harmonic and time dependent wave equations

Participants: Clément Bénéteau, Sonia Fliss.

We study the wave equation set in a periodic half-space when the period is small compared to the wave length. The classical homogenization theory enables to derive an effective model which provides an approximation of the solution. However it is well known that these models are not accurate near the boundaries. In this work, we propose an enriched asymptotic expansion which enables to derive high order effective models at order 1 and 2. Let us mention that the model of order 2 is particularly relevant when one is interested in the long time behaviour of the solution of the time-dependent wave equation. Indeed, it is well-known that the classical homogenized model does not capture the long time dispersion of the exact solution. In several works, homogenized models of order 2 are proposed for the wave equation in infinite domains. Dealing with boundaries and proposing boundary conditions for these models of order 2 were open questions. Our approach enables to propose appropriate and accurate boundary conditions for these models. This work is the fruit of a long time collaboration with Xavier Claeys (LJLL, Sorbonne University) and a recent one with Timothée Pouchon (EPFL).

6.1.2. Interface homogenization

Participant: Jean-François Mercier.

In collaboration with Agnès Maurel from Institut Langevin and Kim Pham from the Department of Mechanics at ENSTA, we have developped interface effective models to describe acoustics and electromagnetic propagation through a scatterers array. The effective models are based on matched asymptotic expansions to account for the small thickness of the array. They consist of determined interface parameters involved in jump conditions for the fields.

1- In acoustics

• Perfect absorption using sparse arrays of Helmholtz resonators

Thanks to an effective model derived to describe a periodic arrangement of Helmholtz resonators, the influence of the spacing on the resonance has been inspected. The strength of the resonance is found enhanced when the array becomes sparser, which provides a degree of freedom to control the radiative damping of the array without affecting the losses within each resonator. It has been used to design a perfect absorbing wall.

• Scattering by arrays of open ended resonators

The previous study has been extended to cavities open at both ends. The effective model provides explicit expressions of the reflection and transmission coefficients, used to provide the relations required to produce zero reflection situation.

• Effective transmission conditions across a resonant bubbly metascreen

The extension to resonant obstacles has been considered with the study of the acoustic propagation through a thin bubbly screen. The analysis is conducted in the time domain and preserves the non linear response of the bubbles. It provides an effective model involving a jump of the normal velocity coupled to an equation of the Rayleigh-Plesset's type for the bubble radius.

2- In electromagnetism

• Perfect Brewster transmission by ultrathin perforated films

The scattering properties of an ultrathin perforated film, made of a material dielectric or perfectly conducting have been studied. Thanks to an asymptotic interface model, the Brewster incidence realizing perfect transmission is accurately described and is found to be significantly shifted from its classical value when the thickness of the film becomes subwavelength.

• Effective transmission conditions for an array of locally resonant inclusions

The previous study has been extended to resonant inclusions of the Mie type. Among the interface parameters involved in the effective model, one is frequency dependent and encapsulates the resonant behavior of the inclusions. Our effective model is validated by comparison with results of full wave calculations.

6.1.3. Wave equation in a weakly randomly perturbed periodic medium

Participants: Sonia Fliss, Laure Giovangigli.

The aim of this work, which is at its first stage, is to construct numerical approximations of the solution of the wave equation in weakly randomly perturbed periodic media in order to propose transparent boundary conditions. We start by studying the effects of rare random perturbations of the medium. The perturbation is weak in the sense that it happens rarely but when it happens the correction is of the order of the initial coefficient. More precisely, we consider the solution of the time harmonic wave equation in a one-dimensional periodic medium, in which each period have a probability η to have its coefficients modified, independently of the other periods. We derive an asymptotic expansion of the distribution of the solution u_η with respect to η and illustrate the convergence with numerical simulations. We also exhibit and implement approximated transparent boundary conditions for such a medium. We then extend the results to more general rare random perturbations. Currently, we are studying other random perturbations of periodic media such as a deformation by a random diffeomorphism with a stationary gradient.

6.1.4. Guided modes in a hexagonal periodic graph-like domain: the zigzag and the armchair cases

Participant: Sonia Fliss.

In this work, we study the wave propagation in hexagonal periodic media that are close to a graph domain. By using an asymptotic analysis, we exhibit situations where the introduction of lineic defects into the geometry of the domain leads to the appearance of guided modes and we show that the direction of the defect leads to very different properties of the guided modes. This work is done in collaboration with Bérangère Delourme (LAGA, Paris 13).

6.1.5. Stable perfectly matched layers for a class of anisotropic dispersive models

Participants: Eliane Bécache, Maryna Kachanovska.

We consider wave propagation in 2D anisotropic dispersive media in an unbounded domain described by Maxwell's equations with an antisymmetric dielectric permittivity tensor and scalar magnetic permeability. Bounding the computational domain is required to obtain the solution. In order to do so, we use the perfectly matched layer (PML) technique. However, the PMLs exhibit instabilities connected to the presence of backward propagating waves. This work is dedicated to stabilizing the PMLs for this case.

6.1.6. Frequency domain wave propagation in anisotropic metamaterials

Participants: Patrick Ciarlet, Maryna Kachanovska.

In this work we address the question of theoretical justification of problems arising in the wave propagation in hyperbolic metamaterials. Such phenomena are described by anisotropic, dispersive Maxwell equations, which, in the frequency domain, correspond to a problem that is hyperbolic for a range of frequencies. For a particular case of such materials (highly magnetized plasmas), we prove the well-posedness of the corresponding model in the free space, providing a suitable radiation condition, as well as study its regularity and demonstrate the limiting amplitude and limiting absorption principles.

6.1.7. On the analysis of perfectly matched layers for electromagnetic waves propagation in anisotropic media

Participants: Eliane Bécache, Anne-Sophie Bonnet-Ben Dhia, Sonia Fliss, Maryna Kachanovska, Maria Kazakova.

This work consists of two parts. The first part is dedicated to the analysis of Cartesian Perfectly Matched Layers (PMLs) in the context of electromagnetic wave propagation in a 3D infinite anisotropic homogeneous medium with a diagonal dielectric tensor. Contrary to the 2D case some anisotropies lead to the existence of backward waves giving rise to instabilities of the PMLs in the time-domain and a lack of convergence in the frequency domain.

The second part examines the behaviour of the PMLs in the frequency domain in the case when in the time domain they give a rise to instabilities. This is the case e.g. for the 2D anisotropic wave equation. For this particular problem, we demonstrate that it is possible to choose the parameters of the PMLs (i.e. the configuration of the PML bounding box and the absorption parameter) to ensure the convergence of the PMLs in the frequency domain.

6.1.8. Maxwell's equations in presence of a conical tip with negative electromagnetic constants

Participants: Anne-Sophie Bonnet-Ben Dhia, Mahran Rihani.

This work is done in collaboration with Lucas Chesnel from CMAP at Ecole Polytechnique. We are interested in the analysis of time-harmonic Maxwell's equations in presence of a conical tip of a material with negative dielectric constants. When these constants belong to some critical range, the electromagnetic field exhibits strongly oscillating singularities at the tip which have infinite energy. In the 2D case of a wedge with critical electromagnetic constants, it has been proved for the equivalent scalar problems that well-posedness in the classical H^1 framework is lost. Well-posedness can be recovered (in a non standard framework) by working in weighted Sobolev spaces and adding in the space the outgoing propagating singularity. We have shown how to provide such functional framework for 3D Maxwell's equations, when only the dielectric permittivity (but not the magnetic permeability) takes a critical value

6.1.9. Essential spectrum related to an interface with a negative material

Participants: Christophe Hazard, Sandrine Paolantoni.

The studies carried out in recent years about the spectral effects of an interface between vacuum and a negative material (that is, a dispersive material whose electric permittivity and magnetic permeability become negative in some frequency range) have been continued in two directions. On the one hand, the previous theoretical studies only considered the non dissipative Drude model. We showed in particular that the interface is responsible for various resonance phenomena related to various components of an essential spectrum. We have extended these results to the so-called Lorentz model (dissipative or not). On the other hand, we have explored the numerical approximation of the spectrum of a cavity partially filled with a Drude material by considering a two-dimensional scalar problem. We have investigated the numerical simulation of the three resonance phenomena associated to the essential spectrum of the cavity.

6.1.10. Computation of plasmon resonances localized at corners using frequency-dependent complex scaling

Participants: Anne-Sophie Bonnet-Ben Dhia, Christophe Hazard, Florian Monteghetti.

A plasmonic device with a non-smooth boundary can exhibit strongly-oscillating surface waves whose phase velocities vanish as they reach the corners. This work investigates in the quasi-static limit the existence of corner resonances, which are analogous to scattering resonances in the sense that the local behavior at each corner plays the role of the behavior at infinity. Resonant contrasts are sought as eigenvalues of the transmission problem with complex scaling applied at corners.

6.1.11. Towards non-local interface models

Participant: Patrick Ciarlet.

Collaboration with Juan Pablo Borthagaray (DMEL, Universidad de la República, Salto, Uruguay). Consider the equation $\mathrm{div}(\sigma\nabla u) = f$ in Ω (plus boundary conditions), where the diffusivity is piecewise constant, and equals σ_i in Ω_i ($i = \{1, 2\}$), with $\overline{\Omega_1} \cup \overline{\Omega_2} = \overline{\Omega}$ and $\Omega_1 \cap \Omega_2 = \varnothing$. If σ_1 and σ_2 have different sign, well-posedness in $H^1(\Omega)$ may not hold. This occurs when the ratio σ_2/σ_1 belongs to the so-called *critical interval*. When the interface has a corner, we have observed that this critical interval is shrunk if one replaces the standard H^1-bilinear forms by corresponding H^s-forms ($s \in (0, 1)$). However, the expense of computing the nonlocal interactions may be prohibitive in applications. Thus, our long term goal is to confine the non-local model to a neighborhood of the interface, while keeping the standard local model in the rest of the domain. A first step in this direction consists in considering the numerical solution of the fractional Laplacian of index $s \in (1/2, 1)$ in a bounded domain Ω with homogeneous Dirichlet boundary conditions. Its solution a priori belongs to the fractional order Sobolev space $\widetilde{H}^s(\Omega)$. Under suitable assumptions on the data, its solution is also in $H^1(\Omega)$. In this case, if one uses the standard Lagrange finite element to discretize the problem, then both the exact and the computed solution belong to $H^1(\Omega)$. We show how to derive error estimates for the Lagrange finite element solutions on both quasi-uniform and graded meshes.

6.1.12. Perturbed edge finite element method for the simulation of electromagnetic waves in magnetised plasmas

Participants: Damien Chicaud, Patrick Ciarlet, Axel Modave.

Numerical simulation of electromagnetic waves in magnetised plasmas is a challenging topic. We address the finite element solution of a time-harmonic model. With the classical method, the variational formulation has a poor coercivity which leads to an ill-conditioned numerical system and numerical instabilities. We propose a perturbed formulation to improve the conditioning of the system. Promising preliminary numerical results have been obtained.

6.1.13. Resonant wave problems in plasmas

Participant: Patrick Ciarlet.

Collaboration with Martin Campos Pinto, Bruno Després and Anouk Nicolopoulos (LJLL, Sorbonne Université). The modelling of resonant waves in 2D plasma leads to the coupling of two degenerate elliptic equations. The model is set over two regions, and involves a smooth, sign-changing coefficient α. The region where $\{\alpha > 0\}$ is propagative, while the region where $\{\alpha < 0\}$ is non propagative, and elliptic. The two models are coupled through the line $\Sigma = \{\alpha = 0\}$. Generically, it is an ill-posed problem, and additional information must be introduced to get a satisfactory treatment at Σ. We define the solution by relying on the limiting absorption principle (the coefficient α is replaced by $\alpha + i0^+$) in an adapted functional setting. This approach relies on the decomposition of the solution in a regular and a singular part, which originates at Σ, and on quasi-solutions. It yields a well-posed mixed variational formulation with coupling. After the design of explicit quasi-solutions, numerical experiments can be carried out, which illustrate the nice properties of this new tool.

6.2. Towards realistic configurations : waveguides and fractals domains

To simulate realistic wave problems, devices which raise specific difficulties concerning either the modeling, the mathematical analysis or the numerical simulation. We start with propagation in *waveguides*, that is a longtime research field within our team, which has acquired an international visibility in this context. We continue with a more recent topic, propagation in *fractal* domains, motivated by a medical application (the human lung).

6.2.1. Transparent boundary conditions for periodic waveguides: analysis and extensions

Participants: Sonia Fliss, Patrick Joly.

We consider the time harmonic wave equation in perturbed periodic waveguides. We justify rigorously the construction of the transparent boundary conditions based on Dirichlet-to-Neumann map and show that the problem with these transparent boundary conditions is of Fredholm type except for a countable set of frequencies. This allows to define and compute the physical solution of the problem. This approach can be applied to deal with junctions of different periodic closed waveguides. We want now to study the extension of the method to the diffraction by locally perturbed periodic layers, surfaces or halfspaces. This work is done in collaboration with Vincent Lescarret (LSS, Centrale Supélec).

6.2.2. Invisible floating objects

Participant: Mahran Rihani.

This work is done in collaboration with Lucas Chesnel from CMAP at Ecole Polytechnique. We consider a time-harmonic water waves problem in a 2D waveguide. The geometry is symmetric with respect to an axis orthogonal to the direction of propagation of waves. Moreover, the waveguide contains two floating obstacles separated by a distance L. We study the behaviours of the scattering coefficients as L goes to ∞. From this analysis, we exhibit situations of non reflectivity or perfect invisibility.

6.2.3. A multi-trace integral equation on infinite boundaries when a global Green's function is not available

Participants: Anne-Sophie Bonnet-Ben Dhia, Sonia Fliss, Yohanes Tjandrawidjaja.

We are interested in time-harmonic scattering problems for configurations where the Green's function is not easily computable for the exterior domain, but different Green's functions are available in several unbounded subdomains covering the whole space. This arises typically for junctions of open waveguides. For a model problem, by using integral representations of the solution in each subdomain, we propose a formulation coupling the traces and the normal traces of the solution on infinite boundaries. The system of equations is shown to have a unique solution in the dissipative case.

6.2.4. Error analysis for transparent boundary conditions in fractal trees

Participants: Patrick Joly, Maryna Kachanovska.

This work is dedicated to an efficient resolution of the wave equation in fractal trees (with application to wave propagation in a human lung). Thanks to self-similarity, it is possible to avoid computing the solution at deeper levels of the tree by using transparent boundary conditions. The corresponding DtN operator is defined by a functional equation for its symbol. in the frequency domain. In this work, we analyse an approximate transparent condition, cf. Waves 2017, based on rational approximation of the symbol. The error and complexity analysis relies on Weyl-like estimates of eigenvalues of the weighted Laplacian and related eigenfunctions.

6.3. Direct and inverse methods for imaging and identification

Imaging and identification, when involved in a real-life context, are often based on wave propagation. This is due to the fact that a substantial part of the information contained in waves can propagate across long distances without significant attenuation. This activity is partly developed in the framework of a long-term partnership with a group of CEA-List in charge of the Non Destructive Testing (NDT) of industrial structures. The aim of NDT is to detect defects inside a structure by imposing some incident waves and measuring the scattered waves caused by the presence of such defects.

6.3.1. The complex-scaled Halfspace Matching Method

Participants: Anne-Sophie Bonnet-Ben Dhia, Christophe Hazard, Sonia Fliss, Yohanes Tjandrawidjaja.

We are currently developing a method that we call the Half-Space Matching (HSM) method, to solve scattering problems in unbounded domains, when classical approaches are either not applicable or too expensive. This method is based on an explicit expression of the "outgoing" solution of the problem in half-spaces, by using Fourier, generalized Fourier or Floquet transforms when the background is respectively homogeneous (possibly anisotropic), stratified or periodic. The domain exterior to a bounded region enclosing the scatterers is recovered by a finite number of halfspaces (at least 3). The unknowns of the formulations are the restriction of the solution to the bounded region and the traces of the solution on the boundary of the halfspaces. The system of equations is derived by writing compatibility conditions between the different representations of the solution. Although the HSM method works in the non-dissipative case, the theoretical and the numerical analysis of the method has been done only in the dissipative case. In the present work, we propose, for the simple case of a homogeneous background, a new formulation of the method which is well-suited for the theoretical and numerical analysis of the non dissipative case. In the spirit of PMLs, the idea is to replace the system of equations on the traces by similar equations on exponentially decaying analytical extensions of the traces.

6.3.2. Implicit-explicit scheme for elastodynamic equations in plates

Participants: Sonia Fliss, Hajer Methenni.

Our objective is to provide an efficient simulation tool for the propagation of elastic waves in thin plates in the context of Guided Waves based Structural Health Monitoring. A naive discretization procedure based on a Leap-frog explicit scheme can be really costly because of the small thickness of the plate. By treating implicitly the operators corresponding to derivatives through the thickness, we show by a stability analysis that the time step is less restricted by the space discretization along the thickness. The price to pay is to solve at each iteration small independent linear systems, but this strategy offers an accurate and efficient discretization of the elastic fields in all dimensions. This method can be used to compute reference solutions and verify the validity of asymptotic models such as Reissner–Mindlin model and some extensions (since there exists no rigorous justifications for elastodynamic problems). Finally under some conditions on the mesh, our approach can be extended to plates with a smoothly varying thickness.

This work is done in collaboration with Sebastien Imperiale (Inria EPI M3DISIM) and Alexandre Imperiale (CEA-LIST).

6.3.3. Forward and inverse scattering in Kirchhoff plates

Participants: Laurent Bourgeois, Christophe Hazard.

A new activity has just started concerning forward and inverse scattering in thin plates governed by the simple Kirchhoff-Love model. The analysis is restricted to the purely bending case and the time-harmonic regime.

We have first considered a 2D strip, that is a waveguide which is unbounded in one direction and bounded in the other (transverse) direction. Two types of conditions on the boundary of the strip are addressed : either the strip is simply supported or the strip is clamped. The two boundary conditions are treated with two different methods. For the simply supported problem, the analysis is based on a result of Hilbert basis in the transverse section. For the clamped problem, this property does not hold. Instead we adopt the Kondratiev's approach, based on the use of the Fourier transform in the unbounded direction, together with techniques of weighted Sobolev spaces with detached asymptotics. After introducing radiation conditions, the corresponding scattering problems in the presence of a free obstacle are shown to be well-posed in the Fredholm sense. We also show that the solutions are the physical (outgoing) solutions in the sense of the limiting absorption principle. This is a joint work Lucas Chesnel, from Inria/DEFI.

We have then addressed the same kind of forward scattering problems for various impenetrable obstacles in an infinite plate. Considering four types of boundary conditions on the obstacle, well-posedness for those problems is proved with the help of a variational approach: (i) for any wave number k when the plate is clamped, simply supported or roller supported; (ii) for any k except a discrete set when the plate is free (this set is finite for convex obstacles). It is then natural to tackle the inverse problem of identifying impenetrable obstacles in a Kirchhoff-Love infinite plate from multistatic near-field data. The Linear Sampling Method is introduced in this context. We firstly prove a uniqueness result for such an inverse problem. We secondly provide the classical theoretical foundation of the Linear Sampling Method. We lastly show the feasibility of the method with the help of numerical experiments. The inverse problem is a joint work with Arnaud Recoquillay, from CEA/LIST.

6.3.4. About regularity and error estimates for the quasi-reversibility method

Participant: Laurent Bourgeois.

This work is done on collaboration with Lucas Chesnel (EPC DEFI). We are interested in the classical ill-posed Cauchy problem for the Laplace equation. One method to approximate the solution associated with compatible data consists in considering a family of regularized well-posed problems depending on a small parameter $\varepsilon > 0$. In this context, in order to prove convergence of finite elements methods, it is necessary to get regularity results of the solutions to these regularized problems which hold uniformly in ε. In the present work, we obtain these results in smooth domains and in 2D polygonal geometries. In the presence of corners, due to the particular structure of the regularized problems, classical techniques *à la* Grisvard do not work and instead, we apply the Kondratiev approach. We describe the procedure in detail to keep track of the dependence in ε in all the estimates. The main originality of this study lies in the fact that the limit problem is ill-posed in any framework.

6.3.5. Analysis of topological derivative as a means for qualitative identification

Participant: Marc Bonnet.

This work is done on collaboration with Fioralba Cakoni, Rutgers University, USA. The concept of topological derivative (TD) has proved effective as a qualitative inversion tool for a wave-based identification of finite-sized objects. Although for the most part, this approach remains based on a heuristic interpretation of the TD, a first attempt toward its mathematical justification was done in Bellis et al. (*Inverse Problems***29**:075012, 2013) for the case of isotropic media with far field data and inhomogeneous refraction index. This work extends the analysis there to the case of anisotropic scatterers and background with near field data. TD-based imaging functional is analyzed using a suitable factorization of the near fields. Our results include justification of sign heuristics for the TD in the isotropic case with jump in the main operator and for some cases of anisotropic media, as well as verifying its decaying property in the isotropic case with near field spherical measurements configuration situated far enough from the probing region.

6.3.6. Asymptotic model for elastodynamic scattering by a small surface-breaking defect

Participant: Marc Bonnet.

This work is done in collaboration with Marc Deschamps and Eric Ducasse, I2M, Bordeaux.

We establish a leading-order asymptotic model for the scattering of elastodynamic fields by small surface-breaking defects in elastic solids. The asymptotic form of the representation formula of the scattered field is written in terms of the elastodynamic Green's tensor, which is in fact available in semi-analytical form for some geometrical configurations that are of practical interest in ultrasonic NDT configurations. Preliminary numerical examples have been performed on cylindrical elastic pipes with small indentations on the outer surface.

6.3.7. Shape optimization of stokesian peristaltic pumps using boundary integral methods

Participant: Marc Bonnet.

This work is done in collaboration with with Ruowen Liu and Shravan Veerapaneni, University of Michigan, USA.

This work develops a new boundary integral approach for finding optimal shapes of peristaltic pumps that transport a viscous fluid. Formulas for computing the shape derivatives of the standard cost functionals and constraints, expressed in boundary-only form, are derived. They involve evaluating physical variables (traction, pressure, etc.) on the boundary only. By employing these formulas in conjunction with a boundary integral approach for solving forward and adjoint problems, we completely avoid the issue of volume remeshing when updating the pump shape as the optimization proceeds. This leads to significant cost savings and we demonstrate the performance on several numerical examples.

6.4. Accelerated numerical solvers for large-scale wave problems

Fast solution procedures are of critical importance for industrial applications such as non-destructive testing, electromagnetic compatibility testing and seismic risk assessment. In these examples, the wavelength is very small in comparison to the characteristic length of the problems, which leads to extremely expensive numerical procedures if standard methods are used. To address the fast numerical solution of large-scale waves problems, we work at the same time on numerical methods, algorithmic issues and implementation strategies to speed up solvers.

6.4.1. Non-overlapping Domain Decomposition Method (DDM) using non-local transmission operators for wave propagation problems.

Participants: Patrick Joly, Emile Parolin.

The research in this direction was mainly concerned by the extension to the electromagnetic setting of the linear convergence theory of non-overlapping DDM that relies on non-local transmission operators. The principal task was to propose, analyse and implement some candidate non-local operators satisfying the assumptions of the theory. There were two main propositions:

- Integral operator for the electromagnetic setting: the operator is available in closed form and its structure lead naturally to a localizable form via truncation of the kernel to limit the effective computational cost while retaining its good properties. The construction of such an operator turned out to be somewhat difficult due to the particular functional setting of Maxwell's equations.
- DtN based non-local operator: the operator is computed by solving auxiliary coercive problems in the vicinity of the transmission interface. The computational cost remains moderate as the implementation no longer involve dense matrix blocks from the integral operators but rather lead to augmented sparse linear systems. Initially developed for the electromagnetic setting, the approach is appealing as it provided a unified formalism that can be applied both to Helmholtz and Maxwell equations and proved to be efficient in numerical experiments.

Another important research direction is created by the technical and theoretical difficulty posed by junction points, which are points where three or more sub-domains abut. Xavier Claeys recently proposed a method to deal with this specific issue, based on the multi-trace formalism, which led to a joint collaboration on the subject. The main idea is to perform a global exchange operation, on the whole skeleton, rather than a local point-to-point exchange. The preliminary numerical results recently obtained are promising.

6.4.2. An efficient domain decomposition method with cross-point treatment for Helmholtz problems

Participant: Axel Modave.

This is a collaboration with X. Antoine (IECL, Nancy), A. Royer (ULiège) and C. Geuzaine (ULiège). The parallel finite-element solution of large-scale time-harmonic scattering problems is addressed with a non-overlapping domain decomposition method (DDM). It is well known that the efficiency of this method strongly depends on the transmission condition enforced on the interfaces between the subdomains. Local conditions based on high-order absorbing boundary conditions (HABCs) are well suited for configurations without cross points *(where more than two subdomains meet)*. In this work, we extend this approach to efficiently deal with cross points. Two-dimensional finite-element results are presented.

6.4.3. Modelling the fluid-structure coupling caused by a far-field underwater explosion using a convolution quadrature based fast boundary element method.

Participants: Marc Bonnet, Stéphanie Chaillat, Damien Mavaleix-Marchessoux.

This study is done in collaboration with Bruno Leblé (Naval Group). It aims at developing computational strategies for modelling the impact of a far-field underwater explosion shock wave on a structure, in deep water. An iterative fluid-structure coupling is developed to solve the problem. Two complementary methods are used: the Finite Element Method (FEM), that offers a wide range of tools to compute the structure response; and the Boundary Element Method (BEM), more suitable to deal with large surrounding fluid domains. We concentrate on developing (i) a fast transient BEM procedure and (ii) a transient FEM-BEM coupling algorithm. The fast transient BEM is based on a fast multipole-accelerated Laplace-domain BEM (implemented in the in-house code COFFEE), extended to the time domain by the Convolution Quadrature Method (CQM). In particular, using empirical approximations for the solution of integral problems involving large (complex) frequencies has been found to yield satisfactorily accurate solutions while saving significant amounts of computational work. The transient BEM-FEM coupling (under progress) will be based on a block-SOR iterative approach, for which a preliminary investigation shows the existence of relaxation parameters that ensure convergence.

6.4.4. Asymptotic based methods for very high frequency problems.

Participant: Eric Lunéville.

This research is developed in collaboration with Marc Lenoir and Daniel Bouche (CEA).

It has recently been realized that the combination of integral and asymptotic methods was a remarkable and necessary tool to solve scattering problems, in the case where the frequency is high and the geometry must be finely taken into account.

In order to implement the high-frequency approximations that we are developing as part of these hybrid HF/BF methods, we have introduced new geometric tools into the XLiFE++ library, in particular splines and B-Splines approximations as well as parameterizations to access quantities such as curvature, curvilinear abscissa, etc. We have also started to interface the OpenCascad library to the XLiFE++ library, which will eventually allow us to manage more complex geometric situations (cylinder and sphere intersection for example). In parallel, we have completed the implementation of 2D HF approximations in the shadow-light transition zone based on the Fock function. Diffraction by a 2D corner is in progress.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Contract and CIFRE PhD with Naval Group on *modelling the fluid-structure coupling caused by a far-field underwater explosion*
  Participants: M. Bonnet, S. Chaillat, D. Mavaleix-Marchessoux
  Start: 11/2017. End: 10/2020. Administrator: CNRS
- Contract and CIFRE PhD with Naval Group on *flow noise prediction*
  Participants: J-F Mercier, B. Cotté, N. Trafny
  Start: 04/2018. End: 03/2021. Administrator: ENSTA
- Contract and CIFRE PhD with CEA on *Modelling of thin layers of randomly distributed nanoparticles for electromagnetic waves* Participants: A. Boucart, S. Fliss, L. Giovangigli
  Start: 10/2019. End: 09/2022. Administrator: ENSTA

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

- ANR project NonlocalDD *(Non-local domain decomposition methods in electromagnetics)*
  Partners: Inria Alpines, Inria POEMS, Inria Magique 3D
  Start: 10/2015. End: 09/2020. Administrator: Inria
  Participants of POEMS: S. Chaillat, P. Joly
  Coordinator: X. Claeys (LJLL, EPI ALPINES)
- ANR project MODULATE *(Modeling lOng-perioD groUnd motions, and assessment of their effects on Large-scale infrAsTructurEs)*
  Partners: ENSTA (UME), Inria POEMS, CentraleSupelec, BRGM, GDS
  Start: 11/2018. End: 10/2021. Administrator: ENSTA
  Participant of POEMS: S. Chaillat
  Coordinator: K. Meza Fajardo (BRGM)

8.1.2. DGA

- Contracts between DGA and POEMS:

- Contract on *boundary element methods and high-frequency problems*
  Participants: E. Lunéville, M. Lenoir, N. Kielbasiewicz.
  Start: 10/2018. End: 09/2021. Administrator: ENSTA
  In partnership with F. Alouges and M. Aussal (CMAP, Ecole Polytechnique).

- DGA provides partial funding for several PhD students:
  - C. Bénéteau on the *asymptotic analysis of time harmonic Maxwell equations in presence of metamaterials* (Start: 10/2017)
  - D. Chicaud on *domain decomposition methods for time-harmonic electromagnetic wave problems with complex media* (Start: 10/2018)

8.2. International Initiatives

8.2.1. Inria International Partners

8.2.1.1. Informal International Partners

Juan Pablo Borthagaray (Universidad de la República, Uruguay)
Shravan Veerapaneni (Univ. of Michigan at Ann Arbor, USA)
Bojan Guzina (University of Minnesota, USA)
Jean-François Molinari (EPFL, Lausanne, Switzerland)
Fioralba Cakoni (University of Rutgers, USA)
Wilkins Aquino (Duke University, USA)
Bojan Guzina (University of Minnesota, USA)
Jorge Albella (University of Santiago de Compostela, Spain)
Carlos Perez Arancibia (Pontificia Universidad Católica, Chile)
Camille Carvalho (UC Merced, Merced, USA)
Simon Chandler Wilde (University of Reading, UK)
Mahadevan Ganesh (Colorado School of Mines, USA)
Christophe Geuzaine (Université de Liège, Belgium)
Marcus Grote (Universitaet Basel, Switzerland)
Moez Khenissi (Univesité de Sousse, Tunisia)
Sergei Nazarov (Saint-Petersburg University, Russia)
Karl-Mikael Perfekt (University of Reading, UK)
Jerónimo Rodríguez (University of Santiago de Compostela, Spain)
Ruben Rosales (MIT, USA)
Adrien Semin (TU Darmstadt, Germany)
Knut Sølna (University of California, Irvine, USA)
Catalin C. Turc (NJIT, NJ, USA)
Jun Zou (Chinese University of Hong Kong, HK)

8.3. International Research Visitors

8.3.1. Visits of International Scientists

- Mahadevan Ganesh (Colorado School of Mines, USA) — March 2019
- Carlos Jerez-Hanckes (Universidad Adolfo Ibanez, Chile) — Septembre 2019
- Shravan Veerapaneni (Univ. of Michigan at Ann Arbor, USA) — November 2019

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. General Chair, Scientific Chair

- Several members of POEMS team were members of the scientific committee of the WAVES conference, organized in Vienna in August 2019.
- J.-F. Mercier co-organized a summer school on *"Wave propagation in complex and microstructured media"* at Cargèse (Corsica), August 20th to 30th, 2019
- A.-S. Bonnet-Ben Dhia co-organized a workshop entitled *“Advanced Theoretical and Numerical Methods for waves in structured Media”* in Marseille in June 2019, in the framework of the GDR Ondes. They were about 90 attendees.

9.1.1.2. Member of the Organizing Committees

- S. Chaillat is a co-animator of the topic *“Modeling and simulation”* of the GDR Ondes.
- J.-F. Mercier is a co-animator of the topic *“Effective dynamics of microstructured media”* of the GDR MecaWaves.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- A. S. Bonnet-Ben Dhia is an associate editor of SIAP *(SIAM Journal of Applied Mathematics)*.
- M. Bonnet is an associate editor of *Engineering Analysis with Boundary Elements*, *Journal of Optimization Theory* and *Application and Journal of Integral Equations and Applications*. He is in the editorial board of *Inverse Problems* and *Computational Mechanics*.
- P. Ciarlet is an editor of ESAIM:M2AN *(Mathematical Modeling and Numerical Analysis)*.
- P. Joly is a member of the Book Series Scientific Computing of Springer Verlag.

9.1.2.2. Reviewer - Reviewing Activities

The team members regularly review papers for many international journals.

9.1.3. Research Administration

- E. Lunéville was chair of the *Applied Mathematics Department* (UMA) at ENSTA Paris until September 2019.
- A. S. Bonnet-Ben Dhia is deputy-chair of the *Applied Mathematics Department* (UMA) at ENSTA Paris since October 2019.
- P. Ciarlet is coordinator of the *Mathematics in Computational Science and Engineering Program* of the Mathematics Hadamard Labex (LMH).
- E. Bécache is a deputy chair of the Doctoral School EDMH since December 2019.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

The permanent members of POEMS are involved in the engineering program at ENSTA Paris, the master program *"Analyse, Modélisation et Simulation"* (AMS) and the master program *"Modélisation et Simulation en Mécanique des Structures et Systèmes Couplés"* (MS2SC) of Université Paris-Saclay.

Eliane Bécache

- *Fonctions de variable complexe*, ENSTA (1st year)

- *Introduction à la discrétisation des équations aux dérivées partielles*, ENSTA (1st year)
- *La méthode des éléments finis*, ENSTA (2nd year) and Master AMS (M1)
- *Analyse et approximation par éléments finis d'EDP*, ENSTA (2nd year) and Master AMS (M1)
- *Résolution des problèmes de diffraction par équations intégrales*, ENSTA (3rd year), Master AMS (M2) and Master MS2SC (M2)

Marc Bonnet

- *Problème inverses et Identification*, Master MS2SC (M2)

Anne-Sophie Bonnet-Ben Dhia

- *Fonctions de variable complexe*, ENSTA (1st year)
- *Théorie spectrale des opérateurs autoadjoints*, ENSTA (2nd year) and Master AMS (M1)
- *Propagation et diffraction dans les guides d'ondes*, ENSTA (3rd year) and Master AMS (M2)
- *Non Destructive Testing*, Master "Acoustical Engineering" (M2)
- *Propagation des ondes élastiques dans les solides*, Master MS2SC (M2)

Laurent Bourgeois

- *Outils élémentaires pour l'analyse des équations aux dérivées partielles*, ENSTA (1st year)
- *Fonctions de variable complexe*, ENSTA (1st year)
- *Complétion de données et identification dans les problèmes gouvernés par des équations aux dérivées partielles*, ENSTA (3rd year) and Master AMS (M2)

Stéphanie Chaillat

- *Introduction à l'environnement UNIX*, ENSTA (1st year)
- *Systèmes d'exploitation*, ENSTA (1st year)
- *Introduction à la discrétisation des équations aux dérivées partielles*, ENSTA (1st year)
- *Résolution des problèmes de diffraction par équations intégrales*, ENSTA (3rd year), Master AMS (M2) and Master MS2SC (M2)
- *Equations intégrales et multipôles rapides*, Ecole doctorale MODES (Univ. Paris Est, Marne la Vallée)

Colin Chambeyron

- *Remise à niveau en maths*, Licence (1st year), Paris-Dauphine University
- *Outils mathématiques*, Licence (L1), Paris-Dauphine University
- *Analyse - Optimisation*, Licence (L1), Paris-Dauphine University
- *Algèbre linéaire*, Licence (L2), Paris-Dauphine University

Patrick Ciarlet

- *Analyse et approximation par éléments finis d'EDP*, ENSTA (2nd year) and Master AMS (M1)
- *Préformation filière ModSim*, ENSTA (3rd year)
- *Modèles mathématiques et leur discrétisation en électromagnétisme*, ENSTA (3rd year) and Master AMS (M2)

Luiz Faria

- *La méthode des éléments finis*, ENSTA (2nd year) and Master AMS (M1)

Sonia Fliss

- *La méthode des éléments finis*, ENSTA (2nd year) and Master AMS (M1)

- *Introduction à la discrétisation des équations aux dérivées partielles*, ENSTA (1st year)
- *Homogénéisation périodique*, ENSTA (3rd year), Master AMS(M2), Masters ANEDP, M4S
- *Propagation des ondes dans des milieux périodiques*, ENSTA (3rd year) and Master AMS (M2)

Laure Giovangigli

- *Introduction aux probabilités et aux statistiques*, ENSTA (1st year)
- *Martingales et algorithmes stochastiques*, ENSTA (2nd year)
- *Calcul stochastique*, ENSTA (3rd year) and Master MMMEF (M2)

Christophe Hazard

- *Outils élémentaires d'analyse pour les équations aux dérivées partielles*, ENSTA (1st year)
- *Théorie spectrale des opérateurs autoadjoints*, ENSTA (2nd year) and Master AMS (M1)

Patrick Joly

- *Introduction à la discrétisation des équations aux dérivées partielles*, ENSTA (1st year)
- *Fonctions de variable complexe*, ENSTA (1st year)
- *Analyse fonctionnelle*, ENSTA (2nd year) and Master AMS (M2)
- *Propagation des ondes dans des milieux périodiques*, ENSTA (3rd year) and Master AMS (M2)

Maryna Kachanovska

- *Fonctions de variable complexe*, ENSTA (1st year)
- *Analyse fonctionnelle*, ENSTA (2nd year) and Master AMS (M1)
- *Modèles mathématiques et leur discrétisation en électromagnétisme*, ENSTA (3rd year) and Master AMS (M2)

Nicolas Kielbasiewicz

- *Programmation scientifique en C++*, ENSTA (2nd year) and Master AMS (M1)
- *Projet de simulation numérique*, ENSTA (2nd year) and Master AMS (M1)
- *Calcul scientifique parallèle*, ENSTA (3rd year) and Master AMS (M2)

Eric Lunéville

- *Introduction au calcul scientifique*, ENSTA (2nd year).
- *Programmation scientifique en C++*, ENSTA (2nd year) and Master AMS (M1)
- *Projet de simulation numérique*, ENSTA (2nd year) and Master AMS (M1)
- *Propagation et diffraction dans les guides d'ondes*, ENSTA (3rd year) and Master AMS (M2)

Jean-François Mercier

- *Outils élémentaires d'analyse pour les équations aux dérivées partielles*, ENSTA (1st year)
- *Fonctions de variable complexe*, ENSTA (1st year)
- *Théorie spectrale des opérateurs autoadjoints*, ENSTA (2nd year) and Master AMS (M1)

Axel Modave

- *Calcul scientifique à haute performance*, ENSTA (2rd year) and Master AMS (M1)
- *Calcul scientifique parallèle*, ENSTA (3rd year) and Master AMS (M2)

9.2.2. Supervision

PhD: Yohanes Tjandrawidjaja, *"Quelques contributions à l'analyse de la Half-Space Matching Method pour les problèmes de diffraction et son extension aux plaques 3D élastiques"*, December 2019, A.-S. Bonnet-Ben Dhia, S. Fliss and V. Baronian

PhD in progress: Sandrine Paolantoni, *"Analyse spectrale et simulation numérique de la diffraction électromagnétique par des métamatériaux"*, October 2016, C. Hazard

PhD in progress: Emile Parolin, *"Non overlapping domain decomposition methods with non local transmission conditions for electromagnetic wave propagation"*, October 2017, P. Joly and X. Claeys

PhD in progress: Clément Bénéteau, *"Asymptotic analysis of time harmonic Maxwell equations in presence of metamaterials"*, October 2017, S. Fliss and X. Claeys

PhD in progress: Hajer Methenni, *"Mathematical modelling and numerical method for the simulation of ultrasound structural health monitoring of composite plates"*, October 2017, S. Fliss and S. Impériale

PhD in progress: Damien Mavaleix-Marchessoux, *"Modeling of the fluid-structure interaction resulting from a remote underwater explosion"*, December 2017, M. Bonnet and S. Chaillat

PhD in progress: Nicolas Trafny, *"Development of semi-analytical models to predict the noise produced by turbulence-edges interactions"*, April 2018, J.-F. Mercier and B. Cotté

PhD in progress: Damien Chicaud, *"Méthodes de décomposition de domaine pour la résolution de problèmes harmoniques d'ondes électromagnétiques en milieux complexes"*, October 2018, P. Ciarlet and A. Modave

PhD in progress: Mahran Rihani, *"Équations de Maxwell en présence de méta-matériaux"*, November 2018, A.-S. Bonnet-Ben Dhia and L. Chesnel

PhD in progress: Akram Beni Hamad, *"Propagation d'ondes électromagnétiques dans les cables coaxiaux"*, Septembre 2019, S. Imperiale, P. Joly and M. Khenissi

PhD in progress: Jean-François Fritsch, *"Imagerie dans les guides d'ondes enfouis"*, Octobre 2019, L. Bourgeois and C. Hazard

PhD in progress: Amandine Boucart *"Modélisation d'une couche mince de nanoparticules réparties aléatoirement pour les ondes électromagnétiques*, Octobre 2019, S. Fliss and L. Giovangigli

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

- A.-S. Bonnet-Ben Dhia is a member of the *bureau du comité des équipes-projets* (BCEP)
- P. Joly is a member of the *commission consultative paritaire scientifique* of Inria
- M. Kachanovska is a member of the *comité scientifique* of Inria-Saclay

9.3.2. Education

- Permanent members of POEMS are involved in the management of the engineering program at ENSTA Paris, the program in applied mathematics at IP Paris, the master 2 program *"Analyse, Modélisation et Simulation"* (M1 AMS) and the master 2 program *"Modélisation et Simulation en Mécanique des Structures et Systèmes Couplés"* (M2 MS2SC) of Université Paris-Saclay:
  - M. Bonnet: co-chair of the M2 MS2SC;
  - L. Bourgeois: co-chair 1A ENSTA until August 2019; co-responsible 2A ENSTA since Septembre 2019; co-chair of the M1 in applied mathematics since Septembre 2019;
  - P. Ciarlet: co-chair 3A ENSTA; deputy head of the M2 AMS until August 2019; coordinator of the master program in applied mathematics at IP Paris;

– S. Fliss: co-chair of the M2 AMS since septembre 2019.

9.3.3. Interventions

- P. Joly presented a talk entitled *"Mathématiques autour de la musique et d'un piano"* during a scientific session at the *Musée des arts et métiers* (March 12, 2019 at Paris) and at the IECL (October 17, 2019 at Nancy).

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] S. CHAILLAT.*Contributions to the modelling of acoustic and elastic wave propagation in large-scale domains with boundary element methods*, ENS Paris Saclay, March 2019, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-02090861

[2] S. FLISS.*Wave propagation in periodic media : mathematical analysis and numerical simulation*, Université Paris Sud (Paris 11), January 2019, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-02394976

Articles in International Peer-Reviewed Journal

[3] F. AMLANI, S. CHAILLAT, A. LOSEILLE.*An efficient preconditioner for adaptive Fast Multipole accelerated Boundary Element Methods to model time-harmonic 3D wave propagation*, in "Computer Methods in Applied Mechanics and Engineering", 2019, vol. 352, n° 1, p. 189-210 [*DOI :* 10.1016/J.CMA.2019.04.026], https://hal.archives-ouvertes.fr/hal-02113702

[4] W. AQUINO, M. BONNET.*Analysis of the error in constitutive equation approach for time-harmonic elasticity imaging*, in "SIAM Journal on Applied Mathematics", 2019, vol. 79, p. 822-849 [*DOI :* 10.1137/18M1231237], https://hal.archives-ouvertes.fr/hal-01948668

[5] D. H. BAFFET, M. J. GROTE, S. IMPERIALE, M. KACHANOVSKA.*Energy Decay and Stability of a Perfectly Matched Layer For the Wave Equation*, in "Journal of Scientific Computing", November 2019 [*DOI :* 10.1007/S10915-019-01089-9], https://hal.archives-ouvertes.fr/hal-01865484

[6] M. BONNET, F. CAKONI.*Analysis of topological derivative as a tool for qualitative identification*, in "Inverse Problems", 2019, vol. 35, n° 104007, https://arxiv.org/abs/1811.10538 [*DOI :* 10.1088/1361-6420/AB0B67], https://hal.archives-ouvertes.fr/hal-01929050

[7] M. BONNET, E. DEMALDENT.*The eddy current model as a low-frequency, high-conductivity asymptotic form of the Maxwell transmission problem*, in "Computers and Mathematics with Applications", 2019, vol. 77, n° 8, p. 2145-2161 [*DOI :* 10.1016/J.CAMWA.2018.12.006], https://hal.archives-ouvertes.fr/hal-01956433

[8] J. P. BORTHAGARAY, P. CIARLET.*On the convergence in H^1-norm for the fractional Laplacian*, in "SIAM Journal on Numerical Analysis", 2019, vol. 57, p. 1723-1743, https://hal.inria.fr/hal-01912092

[9] L. BOURGEOIS, L. CHESNEL.*On quasi-reversibility solutions to the Cauchy problem for the Laplace equation: regularity and error estimates*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2019, forthcoming [*DOI :* 10.1051/M2AN/2019073], https://hal.inria.fr/hal-02385487

[10] L. BOURGEOIS, L. CHESNEL, S. FLISS.*On well-posedness of time-harmonic problems in an unbounded strip for a thin plate model*, in "Communications in Mathematical Sciences", December 2019, https://hal.archives-ouvertes.fr/hal-01883560

[11] L. BOURGEOIS, D. PONOMAREV, J. DARDÉ.*An inverse obstacle problem for the wave equation in a finite time domain*, in "Inverse Problems and Imaging ", April 2019, vol. 19, n^{o} 2, p. 377-400 [*DOI :* 10.3934/IPI.2019019], https://hal.archives-ouvertes.fr/hal-01818956

[12] A. DANSOU, S. MOUHOUBI, C. CHAZALLON, M. BONNET.*Modeling multicrack propagation by the fast multipole symmetric Galerkin BEM*, in "Engineering Analysis with Boundary Elements", September 2019, vol. 106, p. 309-319 [*DOI :* 10.1016/J.ENGANABOUND.2019.05.019], https://hal.archives-ouvertes.fr/hal-02145098

[13] B. DELOURME, S. FLISS, P. JOLY, E. VASILEVSKAYA.*Trapped modes in thin and infinite ladder like domains. Part 2 : asymptotic analysis and numerical application*, in "Asymptotic Analysis", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-01822437

[14] L. FRÉROT, M. BONNET, J.-F. MOLINARI, G. ANCIAUX.*A Fourier-accelerated volume integral method for elastoplastic contact*, in "Computer Methods in Applied Mechanics and Engineering", 2019, vol. 351, p. 951-976, https://arxiv.org/abs/1811.11558 [*DOI :* 10.1016/J.CMA.2019.04.006], https://hal.archives-ouvertes.fr/hal-01963034

[15] I. LABARCA, L. FARIA, C. PÉREZ-ARANCIBIA.*Convolution quadrature methods for time-domain scattering from unbounded penetrable interfaces*, in "Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences", July 2019, vol. 475, n^{o} 2227, 20190029 [*DOI :* 10.1098/RSPA.2019.0029], https://hal.archives-ouvertes.fr/hal-02429472

[16] A. MAUREL, J.-F. MERCIER, T. K. PHAM, J.-J. MARIGO, A. OURIR.*Enhanced resonance of sparse arrays of Helmholtz resonators—Application to perfect absorption*, in "Journal of the Acoustical Society of America", April 2019, vol. 145, n^{o} 4, p. 2552-2560 [*DOI :* 10.1121/1.5098948], https://hal.inria.fr/hal-02381074

[17] A. MODAVE, C. GEUZAINE, X. ANTOINE.*Corner treatments for high-order local absorbing boundary conditions in high-frequency acoustic scattering*, in "Journal of Computational Physics", 2020, vol. 401, 109029 [*DOI :* 10.1016/J.JCP.2019.109029], https://hal.archives-ouvertes.fr/hal-01925160

[18] F. MONTEGHETTI, G. HAINE, D. MATIGNON.*Asymptotic stability of the multidimensional wave equation coupled with classes of positive-real impedance boundary conditions*, in "Mathematical Control and Related Fields", December 2019, vol. 9, n^{o} 4, p. 759-791 [*DOI :* 10.3934/MCRF.2019049], https://hal.archives-ouvertes.fr/hal-02362852

[19] C. PÉREZ-ARANCIBIA, L. FARIA, C. TURC.*Harmonic density interpolation methods for high-order evaluation of Laplace layer potentials in 2D and 3D*, in "Journal of Computational Physics", January 2019, vol. 376, p. 411-434 [*DOI :* 10.1016/J.JCP.2018.10.002], https://hal.archives-ouvertes.fr/hal-02429461

[20] C. PÉREZ-ARANCIBIA, C. TURC, L. FARIA.*Planewave Density Interpolation Methods for 3D Helmholtz Boundary Integral Equations*, in "SIAM Journal on Scientific Computing", July 2019, vol. 41, n^{o} 4, p. A2088-A2116 [*DOI :* 10.1137/19M1239866], https://hal.archives-ouvertes.fr/hal-02429685

Invited Conferences

[21] M. CASSIER, P. JOLY, M. KACHANOVSKA.*Mathematical models for dispersive electromagnetic waves*, in "19w5209 - Herglotz-Nevanlinna Theory Applied to Passive, Causal and Active Systems", Banff International Research Station, Banff, Canada, October 2019, https://hal.archives-ouvertes.fr/hal-02401033

Conferences without Proceedings

[22] G. BECK, M. BONNAUD, J. BENOIT.*Recovering underlying graph for networks of 1D waveguides by reflectometry and transferometry*, in "14th International Conference on Mathematical and Numerical Aspects of Wave Propagation", Vienna, Austria, 2019, https://hal.archives-ouvertes.fr/hal-02414861

[23] G. BECK, S. IMPERIALE, P. JOLY.*Effective models for non-perfectly conducting thin coaxial cables*, in "Waves 2019 - 14th International Conference on Mathematical and Numerical Aspects of Wave Propagation", Vienna, Austria, 2019, https://hal.archives-ouvertes.fr/hal-02414849

[24] A. MODAVE, X. ANTOINE, C. GEUZAINE.*An efficient domain decomposition method with cross-point treatment for Helmholtz problems*, in "CSMA 2019 - 14e Colloque National en Calcul des Structures", Giens (Var), France, May 2019, https://hal.archives-ouvertes.fr/hal-02353415

[25] A. MODAVE, X. ANTOINE, A. ROYER, C. GEUZAINE.*An efficient domain decomposition method with cross-point treatment for Helmholtz problems*, in "WAVES 2019 - 14th International Conference on Mathematical and Numerical Aspects of Wave Propagation", Vienna, Austria, August 2019, https://hal.archives-ouvertes.fr/hal-02353312

Scientific Books (or Scientific Book chapters)

[26] A.-S. BONNET-BEN DHIA, S. FLISS, Y. TJANDRAWIDJAJA.*Numerical analysis of the Half-Space Matching method with Robin traces on a convex polygonal scatterer*, in "Maxwell's equations: Analysis and numerics", Radon Series on Computational and Applied Mathematics, De Gruyter, May 2019, vol. 24, https://hal.inria.fr/hal-01793511

Other Publications

[27] J. ALBELLA MARTÍNEZ, S. IMPERIALE, P. JOLY, J. RODRÍGUEZ.*Numerical Analysis of a Method for Solving 2D Linear Isotropic Elastodynamics with Free Boundary Condition using Potentials and Finite Elements*, November 2019, working paper or preprint [*DOI :* 10.1007/S10915-018-0768-9], https://hal.inria.fr/hal-02345808

[28] L. BOURGEOIS, C. HAZARD.*On well-posedness of scattering problems in a Kirchhoff-Love infinite plate*, October 2019, working paper or preprint, https://hal-ensta-paris.archives-ouvertes.fr//hal-02334004

[29] L. BOURGEOIS, A. RECOQUILLAY.*The Linear Sampling Method for Kirchhoff-Love Infinite Plates*, August 2019, working paper or preprint, https://hal.inria.fr/hal-02269910

[30] R. BUNOIU, L. CHESNEL, K. RAMDANI, M. RIHANI.*Homogenization of Maxwell's equations and related scalar problems with sign-changing coefficients*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02421312

[31] P. CIARLET.*On the approximation of electromagnetic fields by edge finite elements. Part 3: sensitivity to coefficients*, July 2019, working paper or preprint, https://hal.inria.fr/hal-02276430

[32] P. JOLY, M. KACHANOVSKA.*Transparent boundary conditions for wave propagation in fractal trees: convolution quadrature approach (extended report)*, August 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02265345

[33] F. D. KPADONOU, S. CHAILLAT, P. CIARLET.*On the efficiency of nested GMRES preconditioners for 3D acoustic and elastodynamic H-matrix accelerated Boundary Element Methods*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02415902

[34] A. MODAVE, A. ROYER, X. ANTOINE, C. GEUZAINE.*An optimized Schwarz domain decomposition method with cross-point treatment for time-harmonic acoustic scattering*, January 2020, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02432422

[35] A. NICOLOPOULOS, M. CAMPOS PINTO, B. DESPRÉS, P. CIARLET.*Degenerate elliptic equations for resonant wave problems*, July 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02142631

[36] C. PÉREZ-ARANCIBIA, C. TURC, L. FARIA, C. SIDERIS.*Planewave Density Interpolation Methods for the EFIE on Simple and Composite Surfaces*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02429487

Project-Team RANDOPT

Randomized Optimization

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team RANDOPT

Creation of the Team: 2016 December 01, updated into Project-Team: 2019 January 01

Keywords:

Computer Science and Digital Science:

A6.2. - Scientific computing, Numerical Analysis & Optimization
A8.2. - Optimization
A8.9. - Performance evaluation

Other Research Topics and Application Domains:

B4.4. - Energy delivery
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.8. - Reproducibility

1. Team, Visitors, External Collaborators

Research Scientists

Anne Auger [Team leader, Inria, Researcher, HDR]
Dimo Brockhoff [Inria, Researcher]
Nikolaus Hansen [Inria, Senior Researcher, HDR]

PhD Students

Alann Cheral [École polytechnique, PhD Student, from Oct 2019]
Marie Ange Dahito [PSA, PhD Student, granted by CIFRE]
Paul Dufosse [Thales, PhD Student, granted by CIFRE]
Cheikh Saliou Toure [Inria, PhD Student]
Konstantinos Varelas [Thales, PhD Student, granted by CIFRE]
Eugenie Marescaux [Inria, PhD Student, from Sep 2019]

Administrative Assistants

Natalia Alves [Inria, Administrative Assistant, until Jan 2019]
Marie Enee [Inria, Administrative Assistant, from Mar 2019]

2. Overall Objectives

2.1. Scientific Context

Critical problems of the 21st century like the search for highly energy efficient or even carbon-neutral, and cost-efficient systems, or the design of new molecules against extensively drug-resistant bacteria crucially rely on the resolution of challenging numerical optimization problems. Such problems typically depend on noisy experimental data or involve complex numerical simulations such that *derivatives are not useful or not available* and the function is seen as a *black-box*.

Many of those optimization problems are in essence *multiobjective*—one needs to optimize simultaneously several conflicting objectives like minimizing the cost of an energy network and maximizing its reliability—and most of the *challenging* black-box problems are *non-convex, non-smooth* and combine difficulties related to ill-conditioning, non-separability, and ruggedness (a term that characterizes functions that can be non-smooth but also noisy or multi-modal). Additionally the objective function can be expensive to evaluate, that is one function evaluation can take several minutes to hours (it can involve for instance a CFD simulation).

In this context, the use of randomness combined with proper adaptive mechanisms that particularly satisfy several invariance properties (affine invariance, invariance to monotonic transformations) has proven to be one key component for the design of robust global numerical optimization algorithms [35], [25].
The field of adaptive stochastic optimization algorithms has witnessed some important progress over the past 15 years. On the one hand, subdomains like medium-scale unconstrained optimization may be considered as “solved” (particularly, the CMA-ES algorithm, an instance of *Evolution Strategy* (ES) algorithms, stands out as state-of-the-art method) and considerably better standards have been established in the way benchmarking and experimentation are performed. On the other hand, multiobjective population-based stochastic algorithms became the method of choice to address multiobjective problems when a set of some best possible compromises is thought for. In all cases, the resulting algorithms have been naturally transferred to industry (the CMA-ES algorithm is now regularly used in companies such as Bosch, Total, ALSTOM, ...) or to other academic domains where difficult problems need to be solved such as physics, biology [38], geoscience [31], or robotics [33]).

Very recently, ES algorithms attracted quite some attention in Machine Learning with the OpenAI article *Evolution Strategies as a Scalable Alternative to Reinforcement Learning*. It is shown that the training time for difficult reinforcement learning benchmarks could be reduced from 1 day (with standard RL approaches) to 1 hour using ES [36]. [0] A few years ago, another impressive application of CMA-ES, how “Computer Sim Teaches Itself To Walk Upright” (published at the conference SIGGRAPH Asia 2013) was presented in the press in the UK.

Several of those important advances around adaptive stochastic optimization algorithms are relying to a great extent on works initiated or achieved by the founding members of RandOpt particularly related to the CMA-ES algorithm and to the Comparing Continuous Optimizer (COCO) platform.
Yet, the field of adaptive stochastic algorithms for black-box optimization is relatively young compared to the “classical optimization” field that includes convex and gradient-based optimization. For instance, the state-of-the art algorithms for unconstrained gradient based optimization like quasi-Newton methods (e.g. the BFGS method) date from the 1970s [24] while the stochastic derivative-free counterpart, CMA-ES dates from the early 2000s [26]. Consequently, in some subdomains with *important practical demands*, not even the most fundamental and basic questions are answered:

- This is the case of *constrained* optimization where one needs to find a solution $x^* \in \mathbb{R}^n$ minimizing a numerical function $\min_{x \in \mathbb{R}^n} f(x)$ while respecting a number of constraints m typically formulated as $g_i(x^*) \leq 0$ for $i = 1, ..., m$. Only recently, the fundamental requirement of linear convergence [0], as in the unconstrained case, has been clearly stated [16].
- In multiobjective optimization, most of the research so far has been focusing on *how to select candidate solutions from one iteration to the next one*. The difficult question of how to *generate* effectively new solutions is not yet answered in a proper way and we know today that simply applying operators from single-objective optimization may not be effective with the current best selection strategies. As a comparison, in the single-objective case, the question of selection of candidate solutions was already solved in the 1980s and 15 more years were needed to solve the trickier question of an effective adaptive strategy to generate new solutions.

[0] The key behind such an improvement is the parallelization of the algorithm (on thousands of CPUs) that is done in such a way that the communication between the different workers is reduced to only exchanging a vector of permutation of small length (typically less than 100) containing the ranking of candidate solutions on the function to be optimized. In contrast, parallelization of backpropagation requires to exchange the gradient vector of the size of the problem (of the order of 10^6). This reduced communication time is an important factor for the important speedup.

[0] In optimization, linear convergence for an algorithm whose estimate of the optimum x^* of f at iteration t is denoted x_t, refers to a convergence where after a certain time (usually once the initialization is forgotten) the following typically holds: $\|x_{t+1} - x^*\| \leq c\|x_t - x^*\|$ where $c < 1$. This type of convergence is also called geometric. In the case of stochastic algorithms, there exist different definitions of linear convergence (depending on whether we consider the expectation of the sequence or we want a statement that holds with high probability) not strictly equivalent but that always translate the idea that the distance to the optimum at iteration $t + 1$ is a fraction of the distance to the optimum at iteration t.

- With the current demand to solve larger and larger optimization problems (e.g. in the domain of deep learning), optimization algorithms that scale linearly (in terms of internal complexity, memory and number of function evaluations to reach an ϵ-ball around the optimum) with the problem dimension are nowadays needed. Only recently, first proposals of how to reduce the quadratic scaling of CMA-ES have been made without a clear view of what can be achieved in the best case *in practice*. These later variants apply to optimization problems with thousands of variables. The question of designing randomized algorithms capable to handle efficiently problems with one or two orders of magnitude more variables is still largely open.
- For expensive optimization, standard methods are so called Bayesian optimization (BO) algorithms often based on Gaussian processes. Commonly used examples of BO algorithms are EGO [30], SMAC [28], Spearmint [37], or TPE [19] which are implemented in different libraries. Yet, our experience with a popular method like EGO is that many important aspects to come up with a good implementation rely on insider knowledge and are not standard across implementations. Two EGO implementations can differ for example in how they perform the initial design, which bandwidth for the Gaussian kernel is used, or which strategy is taken to optimize the expected improvement.

Additionally, the **development of stochastic adaptive methods for black-box optimization has been mainly driven by heuristics and practice**—rather than a general theoretical framework—validated by intensive computational simulations. Undoubtedly, **this has been an asset as the scope of possibilities for design was not restricted by mathematical frameworks** for proving convergence. In effect, powerful stochastic adaptive algorithms for **unconstrained optimization** like the CMA-ES algorithm emerged from this approach. At the same time, naturally, **theory strongly lags behind practice**. For instance, the striking performances of CMA-ES empirically observed contrast with how little is theoretically proven on the method. This situation is clearly not satisfactory. On the one hand, theory generally lifts performance assessment from an empirical level to a conceptual one, rendering results independent from the problem instances where they have been tested. On the other hand, theory typically provides insights that change perspectives on some algorithm components. Also theoretical guarantees generally increase the trust in the reliability of a method and facilitate the task to make it accepted by wider communities.

Finally, as discussed above, the development of novel black-box algorithms strongly relies on scientific experimentation, and it is quite difficult to conduct proper and meaningful experimental analysis. This is well known for more than two decades now and summarized in this quote from Johnson in 1996

"the field of experimental analysis is fraught with pitfalls. In many ways, the implementation of an algorithm is the easy part. The hard part is successfully using that implementation to produce meaningful and valuable (and publishable!) research results." [29]

Since then, quite some progress has been made to set better standards in conducting scientific experiments and benchmarking. Yet, some domains still suffer from poor benchmarking standards and from the generic problem of the lack of reproducibility of results. For instance, in multiobjective optimization, it is (still) not rare to see comparisons between algorithms made by solely visually inspecting Pareto fronts after a fixed budget. In Bayesian optimization, good performance seems often to be due to insider knowledge not always well described in papers.

2.2. Overall Objectives

In the context of black-box numerical optimization previously described, the scientific positioning of the RandOpt ream is at the intersection between theory, algorithm design, and applications. Our vision is that the field of stochastic black-box optimization should reach the same level of maturity than gradient-based convex mathematical optimization. This entails major algorithmic developments for constrained, multiobjective and large-scale black-box optimization and major theoretical developments for analyzing current methods including the state-of-the-art CMA-ES.

The specificity in black-box optimization is that methods are intended to solve problems characterized by a *non*-property—*non*-convex, *non*-linear, *non*-smooth. This contrasts with gradient-based optimization and poses on the one hand some challenges when developing theoretical frameworks but also makes it compulsory to complement theory with empirical investigations.

Our ultimate goal is to provide software that is useful for practitioners. We see that theory is a means for this end (rather than an end in itself) and it is also our firm belief that parameter tuning is part of the designer's task.

This shapes, on the one hand, four main scientific objectives for our proposed team:

1. **develop novel theoretical frameworks** for guiding (a) the design of novel black-box methods and (b) their analysis, allowing to
2. provide **proofs ofkey features** of stochastic adaptive algorithms including the state-of-the-art method CMA-ES: linear convergence and learning of second order information.
3. develop **stochastic numerical black-box algorithms** following a **principled design** in domains with a strong practical need for much better methods namely **constrained, multiobjective, large-scale and expensive optimization**. Implement the methods such that they are easy to use. And finally, to
4. **set new standards in scientific experimentation, performance assessment and benchmarking** both for optimization on continuous or combinatorial search spaces. This should allow in particular to advance the state of **reproducibility of results of scientific papers** in optimization.

On the other hand, the above motivates our objectives with respect to dissemination and transfer:

1. develop software packages that people can directly use to solve their problems. This means having carefully thought out interfaces, generically applicable setting of parameters and termination conditions, proper treatment of numerical errors, catching properly various exceptions, etc.;
2. have direct collaborations with industrials;
3. publish our results both in applied mathematics and computer science bridging the gap between very often disjoint communities.

3. Research Program

3.1. Introduction

The lines of research we intend to pursue is organized along four axis namely developing novel theoretical framework, developing novel algorithms, setting novel standards in scientific experimentation and benchmarking and applications.

3.2. Developing Novel Theoretical Frameworks for Analyzing and Designing Adaptive Stochastic Algorithms

Stochastic black-box algorithms typically optimize **non-convex, non-smooth functions**. This is possible because the algorithms rely on weak mathematical properties of the underlying functions: the algorithms do not use the derivatives—hence the function does not need to be differentiable—and, additionally, often do not use the exact function value but instead how the objective function ranks candidate solutions (such methods are sometimes called function-value-free).(To illustrate a comparison-based update, consider an algorithm that samples λ (with λ an even integer) candidate solutions from a multivariate normal distribution. Let $x_1, ..., x_\lambda$ in $\mathbb{R}^n$ denote those λ candidate solutions at a given iteration. The solutions are evaluated on the function f to be minimized and ranked from the best to the worse:

$$f(x_{1:\lambda}) \leq ... \leq f(x_{\lambda:\lambda}) \ .$$

In the previous equation $i\!:\!\lambda$ denotes the index of the sampled solution associated to the i-th best solution. The new mean of the Gaussian vector from which new solutions will be sampled at the next iteration can be updated as

$$m \longleftarrow \frac{1}{\lambda} \sum_{i=1}^{\lambda/2} x_{i:\lambda} \ .$$

The previous update moves the mean towards the $\lambda/2$ best solutions. Yet the update is only based on the ranking of the candidate solutions such that the update is the same if f is optimized or $g \circ f$ where $g : \mathrm{Im}(f) \to \mathbb{R}$ is strictly increasing. Consequently, such algorithms are invariant with respect to strictly increasing transformations of the objective function. This entails that they are robust and their performances generalize well.)

Additionally, adaptive stochastic optimization algorithms typically have a **complex state space** which encodes the parameters of a probability distribution (e.g. mean and covariance matrix of a Gaussian vector) and other state vectors. This state-space is a **manifold**. While the algorithms are Markov chains, the complexity of the state-space makes that **standard Markov chain theory tools do not directly apply**. The same holds with tools stemming from stochastic approximation theory or Ordinary Differential Equation (ODE) theory where it is usually assumed that the underlying ODE (obtained by proper averaging and limit for learning rate to zero) has its critical points inside the search space. In contrast, in the cases we are interested in, the **critical points of the ODEs are at the boundary of the domain**.

Last, since we aim at developing theory that on the one hand allows to analyze the main properties of state-of-the-art methods and on the other hand is useful for algorithm design, we need to be careful not to use simplifications that would allow a proof to be done but would not capture the important properties of the algorithms. With that respect one tricky point is to develop **theory that accounts for invariance properties**. To face those specific challenges, we need to develop novel theoretical frameworks exploiting invariance properties and accounting for peculiar state-spaces. Those frameworks should allow researchers to analyze one of the core properties of adaptive stochastic methods, namely **linear convergence** on the widest possible class of functions.

We are planning to approach the question of linear convergence from three different complementary angles, using three different frameworks:

- the Markov chain framework where the convergence derives from the analysis of the stability of a normalized Markov chain existing on scaling-invariant functions for translation and scale-invariant algorithms [18]. This framework allows for a fine analysis where the exact convergence rate can be given as an implicit function of the invariant measure of the normalized Markov chain. Yet it requires the objective function to be scaling-invariant. The stability analysis can be particularly tricky as the Markov chain that needs to be studied writes as $\Phi_{t+1} = F(\Phi_t, W_{t+1})$ where $\{W_t : t > 0\}$ are independent identically distributed and F is typically discontinuous because the algorithms studied are comparison-based. This implies that practical tools for analyzing a standard property like irreducibility, that rely on investigating the stability of underlying deterministic control models [34], cannot be used. Additionally, the construction of a drift to prove ergodicity is particularly delicate when the state space includes a (normalized) covariance matrix as it is the case for analyzing the CMA-ES algorithm.
- The stochastic approximation or ODE framework. Those are standard techniques to prove the convergence of stochastic algorithms when an algorithm can be expressed as a stochastic approximation of the solution of a mean field ODE [20], [21], [32]. What is specific and induces difficulties for the algorithms we aim at analyzing is the **non-standard state-space** since the ODE variables correspond to the state-variables of the algorithm (e.g. $\mathbb{R}^n \times \mathbb{R}_{>0}$ for step-size adaptive algorithms, $\mathbb{R}^n \times \mathbb{R}_{>0} \times S^n_{++}$ where S^n_{++} denotes the set of positive definite matrices if a covariance matrix is additionally adapted). Consequently, the ODE can have many critical points at the boundary of its definition domain (e.g. all points corresponding to $\sigma_t = 0$ are critical points of the ODE) which is

not typical. Also we aim at proving **linear convergence**, for that it is crucial that the learning rate does not decrease to zero which is non-standard in ODE method.

- The direct framework where we construct a global Lyapunov function for the original algorithm from which we deduce bounds on the hitting time to reach an ϵ-ball of the optimum. For this framework as for the ODE framework, we expect that the class of functions where we can prove linear convergence are composite of $g \circ f$ where f is differentiable and $g : \mathrm{Im}(f) \to \mathbb{R}$ is strictly increasing and that we can show convergence to a local minimum.

We expect those frameworks to be complementary in the sense that the assumptions required are different. Typically, the ODE framework should allow for proofs under the assumptions that learning rates are small enough while it is not needed for the Markov chain framework. Hence this latter framework captures better the real dynamics of the algorithm, yet under the assumption of scaling-invariance of the objective functions. Also, we expect some overlap in terms of function classes that can be studied by the different frameworks (typically convex-quadratic functions should be encompassed in the three frameworks). By studying the different frameworks in parallel, we expect to gain synergies and possibly understand what is the most promising approach for solving the holy grail question of the linear convergence of CMA-ES. We foresee for instance that similar approaches like the use of Foster-Lyapunov drift conditions are needed in all the frameworks and that intuition can be gained on how to establish the conditions from one framework to another one.

3.3. Algorithmic developments

We are planning on developing algorithms in the subdomains with strong practical demand for better methods of constrained, multiobjective, large-scale and expensive optimization.

Many of the algorithm developments, we propose, rely on the CMA-ES method. While this seems to restrict our possibilities, we want to emphasize that CMA-ES became a *family of methods* over the years that nowadays include various techniques and developments from the literature to handle non-standard optimization problems (noisy, large-scale, ...). The core idea of all CMA-ES variants—namely the mechanism to adapt a Gaussian distribution—has furthermore been shown to derive naturally from first principles with only minimal assumptions in the context of derivative-free black-box stochastic optimization [35], [25]. This is a strong justification for relying on the CMA-ES premises while new developments naturally include new techniques typically borrowed from other fields. While CMA-ES is now a full family of methods, for visibility reasons, we continue to refer often to “the CMA-ES algorithm”.

3.3.1. Constrained optimization

Many (real-world) optimization problems have constraints related to technical feasibility, cost, etc. Constraints are classically handled in the black-box setting either via rejection of solutions violating the constraints—which can be quite costly and even lead to quasi-infinite loops—or by penalization with respect to the distance to the feasible domain (if this information can be extracted) or with respect to the constraint function value [22]. However, the penalization coefficient is a sensitive parameter that needs to be adapted in order to achieve a robust and general method [23]. Yet, **the question of how to handle properly constraints is largely unsolved**. The latest constraints handling for CMA-ES is an ad-hoc technique driven by many heuristics [23]. Also, it is particularly only recently that it was pointed out that **linear convergence properties should be preserved** when addressing constraint problems [16].

Promising approaches though, rely on using augmented Lagrangians [16], [17]. The augmented Lagrangian, here, is the objective function optimized by the algorithm. Yet, it depends on coefficients that are adapted online. The adaptation of those coefficients is the difficult part: the algorithm should be stable and the adaptation efficient. We believe that the theoretical frameworks developed (particularly the Markov chain framework) will be useful to understand how to design the adaptation mechanisms. Additionally, the question of invariance will also be at the core of the design of the methods: augmented Lagrangian approaches break the invariance to monotonic transformation of the objective functions, yet understanding the maximal invariance that can be achieved seems to be an important step towards understanding what adaptation rules should satisfy.

3.3.2. Large-scale Optimization

In the large-scale setting, we are interested to optimize problems with the order of 10^3 to 10^4 variables. For one to two orders of magnitude more variables, we will talk about a "very large-scale" setting.

In this context, algorithms with a quadratic scaling (internal and in terms of number of function evaluations needed to optimize the problem) cannot be afforded. In CMA-ES-type algorithms, we typically need to restrict the model of the covariance matrix to have only a linear number of parameters to learn such that the algorithms scale linearly in terms of internal complexity, memory and number of function evaluations to solve the problem. The main challenge is thus to have rich enough models for which we can efficiently design proper adaptation mechanisms. Some first large-scale variants of CMA-ES have been derived. They include the online adaptation of the complexity of the model [15], [14]. Yet so far they fail to optimize functions whose Hessian matrix has some small eigenvalues (say around 10^{-4}) some eigenvalues equal to 1 and some very large eigenvalue (say around 10^4), that is functions whose level sets have short and long axis.

Another direction, we want to pursue, is exploring the use of large-scale variants of CMA-ES to solve reinforcement learning problems [36].

Last, we are interested to investigate the very-large-scale setting. One approach consists in doing optimization in subspaces. This entails the efficient identification of relevant spaces and the restriction of the optimization to those subspaces.

3.3.3. Multiobjective Optimization

Multiobjective optimization, i.e., the simultaneous optimization of multiple objective functions, differs from single-objective optimization in particular in its optimization goal. Instead of aiming at converging to the solution with the best possible function value, in multiobjective optimization, a set of solutions [0] is sought. This set, called Pareto-set, contains all trade-off solutions in the sense of Pareto-optimality—no solution exists that is better in *all* objectives than a Pareto-optimal one. Because converging towards a set differs from converging to a single solution, it is no surprise that we might lose many good convergence properties if we directly apply search operators from single-objective methods. However, this is what has typically been done so far in the literature. Indeed, most of the research in stochastic algorithms for multiobjective optimization focused instead on the so called selection part, that decides which solutions should be kept during the optimization—a question that can be considered as solved for many years in the case of single-objective stochastic adaptive methods.

We therefore aim at rethinking search operators and adaptive mechanisms to improve existing methods. We expect that we can obtain orders of magnitude better convergence rates for certain problem types if we choose the right search operators. We typically see two angles of attack: On the one hand, we will study methods based on scalarizing functions that transform the multiobjective problem into a set of single-objective problems. Those single-objective problems can then be solved with state-of-the-art single-objective algorithms. Classical methods for multiobjective optimization fall into this category, but they all solve multiple single-objective problems subsequently (from scratch) instead of dynamically changing the scalarizing function during the search. On the other hand, we will improve on currently available population-based methods such as the first multiobjective versions of the CMA-ES. Here, research is needed on an even more fundamental level such as trying to understand success probabilities observed during an optimization run or how we can introduce non-elitist selection (the state of the art in single-objective stochastic adaptive algorithms) to increase robustness regarding noisy evaluations or multi-modality. The challenge here, compared to single-objective algorithms, is that the quality of a solution is not anymore independent from other sampled solutions, but can potentially depend on all known solutions (in the case of three or more objective functions), resulting in a more noisy evaluation as the relatively simple function-value-based ranking within single-objective optimizers.

3.3.4. Expensive Optimization

In the so-called expensive optimization scenario, a single function evaluation might take several minutes or even hours in a practical setting. Hence, the available budget in terms of number of function evaluation calls

[0] Often, this set forms a manifold of dimension one smaller than the number of objectives.

to find a solution is very limited in practice. To tackle such expensive optimization problems, it is needed to exploit the first few function evaluations in the best way. To this end, typical methods couple the learning of a surrogate (or meta-model) of the expensive objective function with traditional optimization algorithms.

In the context of expensive optimization and CMA-ES, which usually shows its full potential when the number n of variables is not too small (say larger than 3) and if the number of available function evaluations is about $100n$ or larger, several research directions emerge. The two main possibilities to integrate meta-models into the search with CMA-ES type algorithms are (i) the successive injection of the minimum of a learned meta-model at each time step into the learning of CMA-ES's covariance matrix and (ii) the use of a meta-model to predict the internal ranking of solutions. While for the latter, first results exist, the former idea is entirely unexplored for now. In both cases, a fundamental question is which type of meta-model (linear, quadratic, Gaussian Process, ...) is the best choice for a given number of function evaluations (as low as one or two function evaluations) and at which time the type of the meta-model shall be switched.

3.4. Setting novel standards in scientific experimentation and benchmarking

Numerical experimentation is needed as a complement to theory to test novel ideas, hypotheses, the stability of an algorithm, and/or to obtain quantitative estimates. Optimally, theory and experimentation go hand in hand, jointly guiding the understanding of the mechanisms underlying optimization algorithms. Though performing numerical experimentation on optimization algorithms is crucial and a common task, it is non-trivial and easy to fall in (common) pitfalls as stated by J. N. Hooker in his seminal paper [27].

In the RandOpt team we aim at raising the standards for both scientific experimentation and benchmarking.

On the experimentation aspect, we are convinced that there is common ground over how scientific experimentation should be done across many (sub-)domains of optimization, in particular with respect to the visualization of results, testing extreme scenarios (parameter settings, initial conditions, etc.), how to conduct understandable and small experiments, how to account for invariance properties, performing scaling up experiments and so forth. We therefore want to formalize and generalize these ideas in order to make them known to the entire optimization community with the final aim that they become standards for experimental research.

Extensive numerical benchmarking, on the other hand, is a compulsory task for evaluating and comparing the performance of algorithms. It puts algorithms to a standardized test and allows to make recommendations which algorithms should be used preferably in practice. To ease this part of optimization research, we have been developing the Comparing Continuous Optimizers platform (COCO) since 2007 which allows to automatize the tedious task of benchmarking. It is a game changer in the sense that the freed time can now be spent on the scientific part of algorithm design (instead of implementing the experiments, visualization, statistical tests, etc.) and it opened novel perspectives in algorithm testing. COCO implements a thorough, well-documented methodology that is based on the above mentioned general principles for scientific experimentation.

Also due to the freely available data from 300+ algorithms benchmarked with the platform, COCO became a quasi-standard for single-objective, noiseless optimization benchmarking. It is therefore natural to extend the reach of COCO towards other subdomains (particularly constrained optimization, many-objective optimization) which can benefit greatly from an automated benchmarking methodology and standardized tests without (much) effort. This entails particularly the design of novel test suites and rethinking the methodology for measuring performance and more generally evaluating the algorithms. Particularly challenging is the design of scalable non-trivial testbeds for constrained optimization where one can still control where the solutions lies. Other optimization problem types, we are targeting are expensive problems (and the Bayesian optimization community in particular, see our AESOP project), optimization problems in machine learning (for example parameter tuning in reinforcement learning), and the collection of real-world problems from industry.

Another aspect of our future research on benchmarking is to investigate the large amounts of benchmarking data, we collected with COCO during the years. Extracting information about the influence of algorithms on the best performing portfolio, clustering algorithms of similar performance, or the automated detection of anomalies in terms of good/bad behavior of algorithms on a subset of the functions or dimensions are some of the ideas here.

Last, we want to expand the focus of COCO from automatized (large) benchmarking experiments towards everyday experimentation, for example by allowing the user to visually investigate algorithm internals on the fly or by simplifying the set up of algorithm parameter influence studies.

4. Application Domains

4.1. Application Domains

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have a great need to apply black-box algorithms on a daily basis. We do not target a specific application domain and are interested in possible black-box applications stemming from various origins. This is for us intrinsic to the nature of the methods we develop that are general purpose algorithms. Hence our strategy with respect to applications can be seen as opportunistic and our main selection criteria when approached by colleagues who want to develop a collaboration around an application is whether we judge the application interesting: that is the application brings new challenges and/or gives us the opportunity to work on topics we already intended to work on.

The concrete applications related to industrial collaborations we are currently dealing with are:

- With Thales for the theses of Konstantinos Varelas and Paul Dufossé (DGA-CIFRE theses) related to the design of radars (shape optimization of the wave form). Those theses investigate the development of large-scale variants of CMA-ES and constrained-handling for CMA-ES.
- With Storengy, a subsidiary of Engie specialized in gas storage for the thesis of Cheikh Touré. Different multiobjective applications are considered in this context but the primary motivation of Storengy is to get at their disposal a better multiobjective variant of CMA-ES which is the main objective of the developments within the thesis.
- With PSA in the context of the OpenLab and the thesis of Marie-Ange Dahito for the design of part of a car body.
- With Onera in the context of the thesis of Alann Cheral related to the optimization of the choice of hyperspectral bandwidth.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Cheikh Touré received the 2nd prize for the student best paper award at the conference EMO 2019 for the paper “On Bi-Objective convex-quadratic problems” by Cheikh Touré, Anne Auger, Dimo Brockhoff, Nikolaus Hansen
- Nikolaus Hansen received the ENUM best paper award at the ACM-GECCO 2019 conference (see https://gecco-2019.sigevo.org/index.html/Best+Paper+Awards) for the paper “A global Surrogate Assisted CMA-ES”

BEST PAPERS AWARDS :

[6]

C. TOURÉ, A. AUGER, D. BROCKHOFF, N. HANSEN.*On Bi-Objective convex-quadratic problems*, in "10th International Conference on Evolutionary Multi-Criterion Optimization", East Lansing, Michigan, United States, March 2019, https://arxiv.org/abs/1812.00289 , https://hal.inria.fr/hal-01942159

[5]

N. HANSEN.*A Global Surrogate Assisted CMA-ES*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, ACM, 2019, p. 664-672 [*DOI :* 10.1145/3321707.3321842], https://hal.inria.fr/hal-02143961

6. New Software and Platforms

6.1. COCO

COmparing Continuous Optimizers

KEYWORDS: Benchmarking - Numerical optimization - Black-box optimization - Stochastic optimization

SCIENTIFIC DESCRIPTION: COmparing Continuous Optimisers (COCO) is a tool for benchmarking algorithms for black-box optimisation. COCO facilitates systematic experimentation in the field of continuous optimization. COCO provides: (1) an experimental framework for testing the algorithms, (2) post-processing facilities for generating publication quality figures and tables, including the easy integration of data from benchmarking experiments of 300+ algorithm variants, (3) LaTeX templates for scientific articles and HTML overview pages which present the figures and tables.

The COCO software is composed of two parts: (i) an interface available in different programming languages (C/C++, Java, Matlab/Octave, Python, external support for R) which allows to run and log experiments on several function test suites (unbounded noisy and noiseless single-objective functions, unbounded noiseless multiobjective problems, constrained problems) are provided (ii) a Python tool for generating figures and tables that can be looked at in every web browser and that can be used in the provided LaTeX templates to write scientific papers.

FUNCTIONAL DESCRIPTION: The Coco platform aims at supporting the numerical benchmarking of blackbox optimization algorithms in continuous domains. Benchmarking is a vital part of algorithm engineering and a necessary path to recommend algorithms for practical applications. The Coco platform releases algorithm developers and practitioners alike from (re-)writing test functions, logging, and plotting facilities by providing an easy-to-handle interface in several programming languages. The Coco platform has been developed since 2007 and has been used extensively within the “Blackbox Optimization Benchmarking (BBOB)” workshop series since 2009. Overall, 300+ algorithms and algorithm variants by contributors from all over the world have been benchmarked on the platform’s supported test suites so far. The most recent extension has been towards large-scale problems and was used for the BBOB-2019 workshop at the ACM-GECCO conference.

- Participants: Anne Auger, Asma Atamna, Dejan Tusar, Dimo Brockhoff, Marc Schoenauer, Nikolaus HANSEN, Ouassim Ait Elhara, Raymond Ros, Tea Tusar, Thanh-Do Tran, Umut Batu and Konstantinos Varelas
- Partners: TU Dortmund University - Charles University Prague - Jozef Stefan Institute (JSI)
- Contact: Dimo Brockhoff
- URL: https://github.com/numbbo/coco

6.2. CMA-ES

Covariance Matrix Adaptation Evolution Strategy

KEYWORDS: Numerical optimization - Black-box optimization - Stochastic optimization

SCIENTIFIC DESCRIPTION: The CMA-ES is considered as state-of-the-art in evolutionary computation and has been adopted as one of the standard tools for continuous optimisation in many (probably hundreds of) research labs and industrial environments around the world. The CMA-ES is typically applied to unconstrained or bounded constraint optimization problems, and search space dimensions between three and a hundred. The method should be applied, if derivative based methods, e.g. quasi-Newton BFGS or conjugate gradient, (supposedly) fail due to a rugged search landscape (e.g. discontinuities, sharp bends or ridges, noise, local optima, outliers). If second order derivative based methods are successful, they are usually faster than the CMA-ES: on purely convex-quadratic functions, f(x)=xTHx, BFGS (Matlabs function fminunc) is typically faster by a factor of about ten (in terms of number of objective function evaluations needed to reach a target function value, assuming that gradients are not available). On the most simple quadratic function f(x)=||x||2=xTx BFGS is faster by a factor of about 30.

FUNCTIONAL DESCRIPTION: The CMA-ES is an evolutionary algorithm for difficult non-linear non-convex black-box optimisation problems in continuous domain.

- Participant: Nikolaus HANSEN
- Contact: Nikolaus HANSEN
- URL: http://cma.gforge.inria.fr/cmaes_sourcecode_page.html

7. New Results

7.1. A multiobjective algorithm framework and the COMO-CMA-ES algorithm

One classical way to solve multiobjective optimization problems is to transform the multiple objective functions into a single one, also known as scalarization, and to solve multiple versions of such scalarizations to achieve multiple trade-off points. Another approach, coming from the field of evolutionary multiobjective optimization is to formulate a single-objective *set* problem via indicators: the goal here is to find the set of solutions (of a given size) that maximizes a certain quality. The hypervolume indicator has been regularly used to define the quality of the solution set because it has favorable theoretical properties.

The "classical" definition of the hypervolume indicator and how it is used in multiobjective solvers, however, has some disadvantages: (i) the resulting single-objective problem is of high dimension and the gradient of the hypervolume indicator is zero in dominated areas of the search space—not giving a solver enough information about where to search for good solutions. In [7], we discussed and visualized these disadvantages and proposed a new quality criterion which is based on the hypervolume indicator but solves the mentioned disadvantages. The implementation of this idea and its combination with the single-objective solver CMA-ES resulted in the COMO-CMA-ES which shows improved performance over several existing multiobjective solvers on a wide range of convex quadratic functions.

7.2. A mixed-integer benchmark testbed for single and multiobjective black-box optimization

We have introduced two suites of mixed-integer benchmark problems to be used for analyzing and comparing black-box optimization algorithms. They contain problems of diverse difficulties that are scalable in the number of decision variables. The bbob-mixint suite is designed by partially discretizing the established BBOB (Black-Box Optimization Benchmarking) problems. The bi-objective problems from the bbob-biobj-mixint suite are, on the other hand, constructed by using the bbob-mixint functions as their separate objectives. We explain the rationale behind our design decisions and show how to use the suites within the COCO (Comparing Continuous Optimizers) platform. Analyzing two chosen functions in more detail, we also provide some unexpected findings about their properties [8].

7.3. A large-scale optimization testbed for the COCO framework

We have finalized a large scale testbed built to model well-known difficulties in continuous optimization and to test the scaling behavior of algorithms. It contrasts with current test suites used for benchmarking solvers.

The test suite contains 24 single-objective functions in continuous domain and extends the well-known single-objective noiseless bbob test suite. A core contribution is to reduce the computational demand of the orthogonal search space transformations, that appear in the bbob test suite, while retaining some desired properties using previously introduced permuted block diagonal orthogonal matrices.

The paper discusses the implementation details, particularly the normalization and scaling to obtain backwards compatibility with the bbob test suite. Additionally, a guide for using the test suite within the COCO platform, as well as a description of the postprocessed output is presented [12].

7.4. Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies

In [1], we have introduced an acceleration for the covariance matrix adaptation evolution strategies (CMA-ES) by means of adaptive diagonal decoding (dd-CMA). This diagonal acceleration endows the default CMA-ES with the advantages of separable CMA-ES without inheriting its drawbacks. Technically, we introduce a diagonal matrix that expresses coordinate-wise variances of the sampling distribution. The diagonal matrix can learn a rescaling of the problem in the coordinates within linear number of function evaluations. Diagonal decoding can also exploit separability of the problem, but, crucially, does not compromise the performance on non-separable problems. The latter is accomplished by modulating the learning rate for the diagonal matrix based on the condition number of the underlying correlation matrix. dd-CMA-ES not only combines the advantages of default and separable CMA-ES, but may achieve overadditive speedup: it improves the performance, and even the scaling, of the better of default and separable CMA-ES on classes of non-separable test functions that reflect, arguably, a landscape feature commonly observed in practice. The paper makes two further secondary contributions: we introduce two different approaches to guarantee positive definiteness of the covariance matrix with active CMA, which is valuable in particular with large population size; we revise the default parameter setting in CMA-ES, proposing accelerated settings in particular for large dimension. All our contributions can be viewed as independent improvements of CMA-ES, yet they are also complementary and can be seamlessly combined. In numerical experiments with dd-CMA-ES up to dimension 5120, we observe remarkable improvements over the original covariance matrix adaptation on functions with coordinate-wise ill-conditioning. The improvement is observed also for large population sizes up to about dimension squared.

7.5. A global surrogate assisted CMA-ES

In the paper [5], we have explored the arguably simplest way to build an effective surrogate fitness model in continuous search spaces. The model complexity is linear or diagonal-quadratic or full quadratic, depending on the number of available data. The model parameters are computed from the Moore-Penrose pseudoinverse. The model is used as a surrogate fitness for CMA-ES if the rank correlation between true fitness and surrogate value of recently sampled data points is high. Otherwise, further samples from the current population are successively added as data to the model. We empirically compare the IPOP scheme of the new model assisted lq-CMA-ES with a variety of previously proposed methods and with a simple portfolio algorithm using SLSQP and CMA-ES. We conclude that a global quadratic model and a simple portfolio algorithm are viable options to enhance CMA-ES. The model building code is available as part of the pycma Python module on Github and PyPI.

7.6. Benchmarking and Understanding Optimizers

Benchmarking is an important task in optimization in order to understand the working principles behind existing solvers, to find out about weaknesses of them and to finally recommend good ones. The COCO platform, developed in the Randopt team since 2007, aims at automatizing these numerical benchmarking experiments and the visual presentation of their results. We regularly use the platform to initiate workshop papers during the ACM-GECCO conference and also held a workshop this year [0].

In this context, several workshop papers have been published by members of the team and we also proposed some extensions of the platform and updated its documentation.

Two papers addressed single-objective unconstrained problems. One paper investigated the impact of the sample volume of a simple random search on the bbob test suite of COCO [2] and the other paper benchmarked all solvers available in the scipy.optimize module of Python [9], re-discovering SLSQP as a very well-performing solver for small budgets.

Two additional papers addressed multiobjective problems in the context of the bbob-biobj test suite of COCO: “Benchmarking Algorithms from the platypus Framework on the Biobjective bbob-biobj Testbed” [3] compared several baseline algorithms from the literature such as NSGA-II, MOEA/D, SPEA2, and IBEA and “Benchmarking MO-CMA-ES and COMO-CMA-ES on the Bi-objective bbob-biobj Testbed” [4] compared our new COMO-CMA-ES solver with its previous version MO-CMA-ES.

As to extensions of the COCO platform, we released new test suites this year as described earlier. For the large-scale test suite of [12], 11 algorithm variants of the CMA-ES and L-BFGS solvers have been compared in the paper “Benchmarking Large Scale Variants of CMA-ES and L-BFGS-B on the bbob-largescale Testbed” [10].

Overall, we collected 54 new algorithm data sets within the COCO platform in 2019—the highest number in a single year since the release of COCO.

Finally, we updated our documentation on the biobjective test suite, we introduced in 2019. The corresponding journal paper “Using Well-Understood Single-Objective Functions in Multiobjective Black-Box Optimization Test Suites” [11] is now under revision for the Evolutionary Computation journal.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Contract with the company Storengy partially funding the PhD thesis of Cheikh Touré (2017–2020)
- Contract with Thales in the context of the CIFRE PhD thesis of Konstantinos Varelas (2017–2020)
- Contract with PSA in the context of the CIFRE PhD thesis of Marie-Ange Dahito (2019–2022)
- Pending contract for the thesis of Paul Dufossé with Thales (2020–2022)

9. Partnerships and Cooperations

9.1. Regional Initiatives

- PGMO/FMJH project “AESOP: Algorithms for Expensive Simulation-Based Optimization”, 7kEUR, 2017–2019

9.2. National Initiatives

9.2.1. ANR

- ANR project “Big Multiobjective Optimization (BigMO)”, Dimo Brockhoff participates in this project through the Inria team BONUS in Lille (2017–2020)

[0] See numbbo.github.io/workshops/BBOB-2019/

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. Informal International Partners

- Youhei Akimoto, Tsukuba University, Japan
- Tobias Glasmachers, Ruhr University, Bochum, Germany
- Tea Tušar, Jozef Stefan Institute, Ljubljana, Slovenia

9.4. International Research Visitors

9.4.1. Visits to International Teams

9.4.1.1. Research Stays Abroad

Anne Auger and Dimo Brockhoff visited Tea Tušar (Jozef Stefan Institute, Slovenia) for two weeks in April 2019

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Anne Auger: general chair of the ACM-GECCO 2019 conference that took place in July in Prague and welcomed ca. 700 participants.

10.1.1.2. Member of the Organizing Committees

- Dimo Brockhoff: co-organizer of the Lorentz Center Workhop “MACODA: Many Criteria Optimization and Decision Analysis“, September 2019, Leiden, The Netherlands, 42 participants
- Anne Auger, Dimo Brockhoff, Nikolaus Hansen, and Konstantinos Varelas: co-organizers of the Black-Box Optimization Benchmarking workshop (BBOB) at the ACM-GECCO 2019 conference.
- Anne Auger: Organizer of an **invited session** on *Recent Advances on Randomized Derivative Free Optimization* at the EURO conference, Dublin.

10.1.2. Scientific Events: Selection

10.1.2.1. Reviewer

- Dimo Brockhoff and Nikolaus Hansen: ACM-GECCO 2019

10.1.3. Dagstuhl seminar invitations

- Anne Auger, Dimo Brockhoff and Nikolaus Hansen invited at the Dagstuhl Seminar 19431 on “Theory of Randomized Optimization Heuristics”, October 2019
- Dimo Brockhoff invited at the Dagstuhl Seminar 20031 on “Scalability in Multiobjective Optimization”, January 2020

10.1.4. Journal

10.1.4.1. Member of the Editorial Boards

- Anne Auger, Dimo Brockhoff and Nikolaus Hansen: Associate Editor of the ACM Transactions on Evolutionary Learning and Optimization
- Anne Auger and Nikolaus Hansen: Associate Editor of the Evolutionary Computation Journal

- Anne Auger is guest editor of an Algorithmica special issue of papers selected from the ACM-GECCO'2018 theory track
- Anne Auger is guest editor of the IEEE Transactions on Evolutionary Computation special issue on Theoretical Foundations of Evolutionary Computation

10.1.4.2. Reviewer - Reviewing Activities

- Dimo Brockhoff: reviewed papers for IEEE Transactions on Evolutionary Computation (IEEETEC), European Journal of Operational Research (EJOR), Evolutionary Computation, and Journal of Global Optimization

10.1.5. Invited Talks

- Dimo Brockhoff: "Quality Indicator Maximization in Multiobjective Optimization Via Single-Objective Solvers: Unflattened Hypervolume Improvement in the Sofomore Framework", Jozef Stefan Institute, Ljubljana, Slovenia, April 2019
- Nikolaus Hansen: "How to Evolve Gradient Descent into Evolution Strategies and CMA-ES", Journée de Rentrée du CMAP, Ecole polytechnique, Paris, October 2019
- Nikolaus Hansen: "How to Evolve Gradient Descent into Evolution Strategies and CMA-ES", Symposium on Evolutionary Algorithms: Back to the Future and Beyond—Traversing the Ever-Evolving Landscape of Evolutionary Algorithms, Delft, The Netherlands, September 2019
- Nikolaus Hansen: Keynote lecture at the Workshop on Machine-Learning-Assisted Image Formation entitled "From Gradient-Based to Evolutionary Optimization", Nice, France, June 2019
- Anne Auger: Convergence Results of Adaptive Evolution Strategies: an Overview, Lifeware seminar, January 2019
- Anne Auger: COMO-CMA-ES a linearly convergent multi-objective solver, ICCOP conference, Berlin

10.1.6. Leadership within the Scientific Community

- Anne Auger, Elected Member of the ACM-SIGEVO executive board
- Dimo Brockhoff: member of the International Advisory Committee for EMO-2019 in East Lansing, USA
- Anne Auger, member of the conseil de laboratoire of the CMAP, Ecole Polytechnique.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Anne Auger, "Optimization without gradients", 22.5h ETD, niveau M2 (Optimization Master of Paris-Saclay)

Master: Dimo Brockhoff, "Algorithms and Complexity", 36h ETD, niveau M1/M2 (joint MSc with ESSEC "Data Sciences & Business Analytics"), CentraleSupelec, France

Master: Anne Auger and Dimo Brockhoff, "Introduction to Optimization", 31.5h ETD, niveau M2 (MSc Informatique - Parcours Apprentissage, Information et Contenu (AIC)), U. Paris-Saclay, France

Master: Anne Auger and Dimo Brockhoff, "Advanced Optimization", 31.5h ETD, niveau M2 (MSc Informatique - Parcours Apprentissage, Information et Contenu (AIC)), U. Paris-Saclay, France

10.2.2. Tutorials

- Dimo Brockhoff: tutorial on Evolutionary Multiobjective Optimization, 3h ETD, niveau PhD, ACM-GECCO conference, Czech Republic

- Nikolaus Hansen: tutorial “A Practical Guide to Experimentation”, 3h ETD, niveau PhD, ACM-GECCO conference, Czech Republic
- Nikolaus Hansen: tutorial on “CMA-ES and Advanced Adaptation Mechanisms”, 3h ETD, niveau PhD, ACM-GECCO conference, Czech Republic, with Youhei Akimoto

10.2.3. Supervision

- PhD in progress: Konstantinos Varelas, “Large-Scale Optimization, CMA-ES and Radar Applications”, Dec. 2017, Anne Auger and Dimo Brockhoff
- PhD in progress: Cheikh Touré, “Linearly Convergent Multi-objective Stochastic Optimizers”, Dec. 2017, Anne Auger and Dimo Brockhoff
- PhD in progress: Paul Dufossé, “Constrained Optimization and Radar Applications”, Oct. 2018, Nikolaus Hansen
- PhD in progress: Marie-Ange Dahito, “Mixed-Integer Blackbox Optimization for Multiobjective Problems in the Automotive Industry”, Jan 2019, Dimo Brockhoff and Nikolaus Hansen
- PhD in progress: Eugénie Marescaux, Theoretical Analysis of convergence of multi-objective solvers (2019–), supervisor A. Auger
- PhD in progress: Alann Cheral, “Black-box optimization for the optimization of hyperspectral bandwidth for anomaly detection” (2019–), supervisor A. Auger
- Jingyun Yang, Ecole Polytechnique, since September 2019
- Eugenie Marescaux, ENSTA, April–September 2019
- Julien Bonneville, U. Versailles, March–August 2019

10.2.4. Juries

- Anne Auger: member of the **PGMO PhD award** scientific committee. Representing the committee for the PGMO PhD award ceremony at the PGMO days
- Dimo Brockhoff: jury member for the PhD thesis of David Gaudrie, École des Mines de Saint-Étienne, October 2019

10.3. Popularization

- Scientific mediation by Cheikh Touré: presentation on poker in the context of the visit of college and high school students to Inria Saclay

10.3.1. Internal or external Inria responsibilities

- Anne Auger: member of the BCEP of Saclay.
- Dimo Brockhoff: member of the CDT committee of Saclay (since February 2019).

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] Y. AKIMOTO, N. HANSEN.*Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies*, in "Evolutionary Computation", 2019, https://arxiv.org/abs/1905.05885 , forthcoming [*DOI :* 10.1162/EVCO_A_00260], https://hal.inria.fr/hal-01995373

International Conferences with Proceedings

[2] D. BROCKHOFF, N. HANSEN.*The Impact of Sample Volume in Random Search on the bbob Test Suite*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019 [*DOI :* 10.1145/3319619.3326894], https://hal.inria.fr/hal-02171213

[3] D. BROCKHOFF, T. TUŠAR.*Benchmarking Algorithms from the platypus Framework on the Biobjective bbob-biobj Testbed*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019, vol. 7 [*DOI :* 10.1145/3319619.3326896], https://hal.inria.fr/hal-02171136

[4] P. DUFOSSÉ, C. TOURÉ.*Benchmarking MO-CMA-ES and COMO-CMA-ES on the Bi-objective bbob-biobj Testbed*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019 [*DOI :* 10.1145/3319619.3326892], https://hal.inria.fr/hal-02161252

[5] *Best Paper*
N. HANSEN.*A Global Surrogate Assisted CMA-ES*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, ACM, 2019, p. 664-672 [*DOI :* 10.1145/3321707.3321842], https://hal.inria.fr/hal-02143961.

[6] *Best Paper*
C. TOURÉ, A. AUGER, D. BROCKHOFF, N. HANSEN.*On Bi-Objective convex-quadratic problems*, in "10th International Conference on Evolutionary Multi-Criterion Optimization", East Lansing, Michigan, United States, March 2019, https://arxiv.org/abs/1812.00289 , https://hal.inria.fr/hal-01942159.

[7] C. TOURÉ, N. HANSEN, A. AUGER, D. BROCKHOFF.*Uncrowded Hypervolume Improvement: COMO-CMA-ES and the Sofomore framework*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019, Part of this research has been conducted in the context of a research collaboration between Storengy and Inria [*DOI :* 10.1145/3321707.3321852], https://hal.inria.fr/hal-02103694

[8] T. TUŠAR, D. BROCKHOFF, N. HANSEN.*Mixed-Integer Benchmark Problems for Single-and Bi-Objective Optimization*, in "GECCO 2019 -The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019, submitted to GECCO 2019, https://hal.inria.fr/hal-02067932

[9] K. VARELAS, M.-A. DAHITO.*Benchmarking Multivariate Solvers of SciPy on the Noiseless Testbed*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019 [*DOI :* 10.1145/3319619.3326891], https://hal.inria.fr/hal-02160099

[10] K. VARELAS.*Benchmarking Large Scale Variants of CMA-ES and L-BFGS-B on the bbob-largescale Testbed*, in "GECCO 2019 - The Genetic and Evolutionary Computation Conference", Prague, Czech Republic, July 2019 [*DOI :* 10.1145/3319619.3326893], https://hal.inria.fr/hal-02160106

Other Publications

[11] D. BROCKHOFF, T. TUŠAR, A. AUGER, N. HANSEN.*Using Well-Understood Single-Objective Functions in Multiobjective Black-Box Optimization Test Suites*, January 2019, https://arxiv.org/abs/1604.00359 - ArXiv e-prints, arXiv:1604.00359, https://hal.inria.fr/hal-01296987

[12] O. A. ELHARA, K. VARELAS, D. H. NGUYEN, T. TUŠAR, D. BROCKHOFF, N. HANSEN, A. AUGER.*COCO: The Large Scale Black-Box Optimization Benchmarking (bbob-largescale) Test Suite*, March 2019, https://arxiv.org/abs/1903.06396 - working paper or preprint, https://hal.inria.fr/hal-02068407

[13] S. MAHÉVAS, V. PICHENY, P. LAMBERT, N. DUMOULIN, L. ROUAN, C. SOULIÉ, D. BROCKHOFF, S. LEHUTA, R. LE RICHE, R. FAIVRE, H. DROUINEAU.*A practical guide for conducting calibration and decision-making optimisation with complex ecological models*, December 2019, pdf available at https://www.preprints.org/manuscript/201912.0249/v1 [*DOI :* 10.20944/PREPRINTS201912.0249.V1], https://hal.inria.fr/hal-02418667

References in notes

[14] Y. AKIMOTO, N. HANSEN.*Online model selection for restricted covariance matrix adaptation*, in "International Conference on Parallel Problem Solving from Nature", Springer, 2016, p. 3–13

[15] Y. AKIMOTO, N. HANSEN.*Projection-based restricted covariance matrix adaptation for high dimension*, in "Proceedings of the 2016 on Genetic and Evolutionary Computation Conference", ACM, 2016, p. 197–204

[16] D. V. ARNOLD, J. PORTER.*Towards au Augmented Lagrangian Constraint Handling Approach for the* $(1+1)$*-ES*, in "Genetic and Evolutionary Computation Conference", ACM Press, 2015, p. 249-256

[17] A. ATAMNA, A. AUGER, N. HANSEN.*Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling*, in "Foundation of Genetic Algorithms (FOGA)", 2017

[18] A. AUGER, N. HANSEN.*Linear Convergence of Comparison-based Step-size Adaptive Randomized Search via Stability of Markov Chains*, in "SIAM Journal on Optimization", 2016, vol. 26, n^{o} 3, p. 1589-1624

[19] J. BERGSTRA, R. BARDENET, Y. BENGIO, B. KÉGL.*Algorithms for Hyper-Parameter Optimization*, in "Neural Information Processing Systems (NIPS 2011)", 2011, https://hal.inria.fr/hal-00642998/file/draft1.pdf

[20] V. S. BORKAR.*Stochastic approximation: a dynamical systems viewpoint*, 2008, Cambridge University Press

[21] V. BORKAR, S. MEYN.*The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning*, in "SIAM Journal on Control and Optimization", January 2000, vol. 38, n^{o} 2

[22] C. A. COELLO COELLO.*Constraint-handling techniques used with evolutionary algorithms*, in "Proceedings of the 2008 Genetic and Evolutionary Computation Conference", ACM, 2008, p. 2445–2466

[23] G. COLLANGE, S. REYNAUD, N. HANSEN.*Covariance Matrix Adaptation Evolution Strategy for Multidisciplinary Optimization of Expendable Launcher Families*, in "13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, Proceedings", 2010

[24] J. E. DENNIS, R. B. SCHNABEL.*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983

[25] N. HANSEN, A. AUGER.*Principled design of continuous stochastic search: From theory to practice*, in "Theory and principled methods for the design of metaheuristics", Springer, 2014, p. 145–180

[26] N. HANSEN, A. OSTERMEIER.*Completely Derandomized Self-Adaptation in Evolution Strategies*, in "Evolutionary Computation", 2001, vol. 9, n^{o} 2, p. 159–195

[27] J. N. HOOKER.*Testing heuristics: We have it all wrong*, in "Journal of heuristics", 1995, vol. 1, n^{o} 1, p. 33–42

[28] F. HUTTER, H. HOOS, K. LEYTON-BROWN.*An Evaluation of Sequential Model-based Optimization for Expensive Blackbox Functions*, in "GECCO (Companion) 2013", ACM, 2013, p. 1209–1216

[29] D. S. JOHNSON.*A theoretician's guide to the experimental analysis of algorithms*, in "Data structures, near neighbor searches, and methodology: fifth and sixth DIMACS implementation challenges", 2002, vol. 59, p. 215–250

[30] D. R. JONES, M. SCHONLAU, W. J. WELCH.*Efficient global optimization of expensive black-box functions*, in "Journal of Global optimization", 1998, vol. 13, n^{o} 4, p. 455–492

[31] I. KRIEST, V. SAUERLAND, S. KHATIWALA, A. SRIVASTAV, A. OSCHLIES.*Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0)*, in "Geoscientific Model Development", 2017, vol. 10, n^{o} 1, 127

[32] H. J. KUSHNER, G. YIN.*Stochastic approximation and recursive algorithms and applications*, Applications of mathematics, Springer, New York, 2003, http://opac.inria.fr/record=b1099801

[33] P. MACALPINE, S. BARRETT, D. URIELI, V. VU, P. STONE.*Design and Optimization of an Omnidirectional Humanoid Walk: A Winning Approach at the RoboCup 2011 3D Simulation Competition*, in "Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI)", July 2012

[34] S. MEYN, R. TWEEDIE.*Markov Chains and Stochastic Stability*, Springer-Verlag, New York, 1993

[35] Y. OLLIVIER, L. ARNOLD, A. AUGER, N. HANSEN.*Information-geometric optimization algorithms: A unifying picture via invariance principles*, in "Journal Of Machine Learning Research", 2016, accepted

[36] T. SALIMANS, J. HO, X. CHEN, I. SUTSKEVER.*Evolution strategies as a scalable alternative to reinforcement learning*, in "arXiv preprint arXiv:1703.03864", 2017

[37] J. SNOEK, H. LAROCHELLE, R. P. ADAMS.*Practical bayesian optimization of machine learning algorithms*, in "Neural Information Processing Systems (NIPS 2012)", 2012, p. 2951–2959

[38] J. UHLENDORF, A. MIERMONT, T. DELAVEAU, G. CHARVIN, F. FAGES, S. BOTTANI, G. BATT, P. HERSEN.*Long-term model predictive control of gene expression at the population and single-cell levels*, in "Proceedings of the National Academy of Sciences", 2012, vol. 109, n^{o} 35, p. 14271–14276

Project-Team SPECFUN

Symbolic Special Functions : Fast and Certified

RESEARCH CENTER
Saclay - Île-de-France

THEME
Algorithmics, Computer Algebra and Cryptology

Table of contents

Project-Team SPECFUN

Creation of the Team: 2012 November 01, updated into Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:

A2.1.11. - Proof languages
A2.4.3. - Proofs
A4.5. - Formal methods for security
A7.2. - Logic in Computer Science
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology
A8.4. - Computer Algebra
A8.5. - Number theory

Other Research Topics and Application Domains:

B9.5.2. - Mathematics
B9.5.3. - Physics

1. Team, Visitors, External Collaborators

Research Scientists

Frédéric Chyzak [Team leader, Inria, Researcher, HDR]
Alin Bostan [Inria, Researcher, HDR]
Georges Gonthier [Inria, Senior Researcher]
Pierre Lairez [Inria, Researcher]

External Collaborators

Philippe Dumas [Ministère de l'Education Nationale (retired)]
Guy Fayolle [Inria, Senior Researcher (emeritus), part time, also with Project-Team RITS]
Marc Mezzarobba [CNRS]

Post-Doctoral Fellow

Svyatoslav Covanov [Inria, Post-Doctoral Fellow, until Oct 2019]

Visiting Scientist

Antonio Jimenez Pastor [RISC, from May 2019 until Jun 2019]

Administrative Assistant

Bahar Carabetta [Inria, from Feb 2019]

2. Overall Objectives

2.1. Scientific challenges, expected impact

The general orientation of our team is described by the short name given to it: *Special Functions*, that is, particular mathematical functions that have established names due to their importance in mathematical analysis, physics, and other application domains. Indeed, we ambition to study special functions with the computer, by combined means of computer algebra and formal methods.

Computer-algebra systems have been advertised for decades as software for "doing mathematics by computer" [68]. For instance, computer-algebra libraries can uniformly generate a corpus of mathematical properties about special functions, so as to display them on an interactive website. This possibility was recently shown by the computer-algebra component of the team [23]. Such an automated generation significantly increases the reliability of the mathematical corpus, in comparison to the content of existing static authoritative handbooks. The importance of the validity of these contents can be measured by the very wide audience that such handbooks have had, to the point that a book like [18] remains one of the most cited mathematical publications ever and has motivated the 10-year-long project of writing its successor [20]. However, can the mathematics produced "by computer" be considered as *true* mathematics? More specifically, whereas it is nowadays well established that the computer helps in discovering and observing new mathematical phenomenons, can the mathematical statements produced with the aid of the computer and the mathematical results computed by it be accepted as valid mathematics, that is, as having the status of mathematical *proofs*? Beyond the reported weaknesses or controversial design choices of mainstream computer-algebra systems, the issue is more of an epistemological nature. It will not find its solution even in the advent of the ultimate computer-algebra system: the social process of peer-reviewing just falls short of evaluating the results produced by computers, as reported by Th. Hales [47] after the publication of his proof of the Kepler Conjecture about sphere packing.

A natural answer to this deadlock is to move to an alternative kind of mathematical software and to use a proof assistant to check the correctness of the desired properties or formulas. The success of large-scale formalization projects, like the Four-Color Theorem of graph theory [42], the above-mentioned Kepler Conjecture [47], and the Odd Order Theorem of group theory [0], have increased the understanding of the appropriate software-engineering methods for this peculiar kind of programming. For computer algebra, this legitimates a move to proof assistants now.

The Dynamic Dictionary of Mathematical Functions [0] (DDMF) [23] is an online computer-generated handbook of mathematical functions that ambitions to serve as a reference for a broad range of applications. This software was developed by the computer-algebra component of the team as a project [0] of the MSR–INRIA Joint Centre. It bases on a library for the computer-algebra system Maple, Algolib [0], whose development started 20 years ago in project-team Algorithms [0]. As suggested by the constant questioning of certainty by new potential users, DDMF deserves a formal guarantee of correctness of its content, on a level that proof assistants can provide. Fortunately, the maturity of special-functions algorithms in Algolib makes DDMF a stepping stone for such a formalization: it provides a well-understood and unified algorithmic treatment, without which a formal certification would simply be unreachable.

The formal-proofs component of the team emanates from another project of the MSR–INRIA Joint Centre, namely the Mathematical Components project (MathComp) [0]. Since 2006, the MathComp group has endeavoured to develop computer-checked libraries of formalized mathematics, using the Coq proof assistant [64]. The methodological aim of the project was to understand the design methods leading to successful large-scale formalizations. The work culminated in 2012 with the completion of a formal proof of the Odd Order Theorem, resulting in the largest corpus of algebraic theories ever machine-checked with a proof assistant and a whole methodology to effectively combine these components in order to tackle complex formalizations. In particular, these libraries provide a good number of the many algebraic objects needed to reason about special functions and their properties, like rational numbers, iterated sums, polynomials, and a rich hierarchy of algebraic structures.

The present team takes benefit from these recent advances to explore the formal certification of the results collected in DDMF. The aim of this project is to concentrate the formalization effort on this delimited area, building on DDMF and the Algolib library, as well as on the Coq system [64] and on the libraries developed by the MathComp project.

[0] http://www.msr-inria.inria.fr/news/the-formalization-of-the-odd-order-theorem-has-been-completed-the-20-septembre-2012/
[0] http://ddmf.msr-inria.inria.fr/1.9.1/ddmf
[0] http://www.msr-inria.inria.fr/projects/dynamic-dictionary-of-mathematical-functions/
[0] http://algo.inria.fr/libraries/
[0] http://algo.inria.fr/
[0] http://www.msr-inria.fr/projects/mathematical-components/

2.1.1. Use computer algebra but convince users beyond reasonable doubt

The following few opinions on computer algebra are, we believe, typical of computer-algebra users' doubts and difficulties when using computer-algebra systems:

- Fredrik Johansson, expert in the multi-precision numerical evaluation of special functions and in fast computer-algebra algorithms, writes on his blog [53]: "Mathematica is great for cross-checking numerical values, but it's not unusual to run into bugs, so *triple checking is a good habit*." One answer in the discussion is: "We can claim that Mathematica has [...] *an impossible to understand semantics*: If Mathematica's output is wrong then change the input. If you don't like the answer, change the question. That seems to be the philosophy behind."
- A professor's advice to students [60] on using Maple: "You may wish to use Maple to check your homework answers. If you do then keep in mind that Maple sometimes gives the *wrong answer, usually because you asked incorrectly, or because of niceties of analytic continuation*. You may even be bitten by an occasional Maple bug, though that has become fairly unlikely. Even with as powerful a tool as Maple you will still *have to devise your own checks* and you will still have to think."
- Jacques Carette, former head of the maths group at Maplesoft, about a bug [19] when asking Maple to take the limit `limit(f(n) * exp(-n), n = infinity)` for an undetermined function `f`: "The problem is that there is an *implicit assumption in the implementation* that unknown functions do not 'grow too fast'."

As explained by the expert views above, complaints by computer-algebra users are often due to their misunderstanding of what a computer-algebra systems is, namely a purely syntactic tool for calculations, that the user must complement with a semantics. Still, robustness and consistency of computer-algebra systems are not ensured as of today, and, whatever Zeilberger may provocatively say in his Opinion 94 [69], a firmer logical foundation is necessary. Indeed, the fact is that many "bugs" in a computer-algebra system cannot be fixed by just the usual debugging method of tracking down the faulty lines in the code. It is sort of "by design": assumptions that too often remain implicit are really needed by the design of symbolic algorithms and cannot easily be expressed in the programming languages used in computer algebra. A similar certification initiative has already been undertaken in the domain of numerical computing, in a successful manner [51], [26]. It is natural to undertake a similar approach for computer algebra.

2.1.2. Make computer algebra and formal proofs help one another

Some of the mathematical objects that interest our team are still totally untouched by formalization. When implementing them and their theory inside a proof assistant, we have to deal with the pervasive discrepancy between the published literature and the actual implementation of computer-algebra algorithms. Interestingly, this forces us to clarify our computer-algebraic view on them, and possibly make us discover holes lurking in published (human) proofs. We are therefore convinced that the close interaction of researchers from both fields, which is what we strive to maintain in this team, is a strong asset.

For a concrete example, the core of Zeilberger's creative telescoping manipulates rational functions up to simplifications. In summation applications, checking that these simplifications do not hide problematic divisions by 0 is most often left to the reader. In the same vein, in the case of integrals, the published algorithms do not check the convergence of all integrals, especially in intermediate calculations. Such checks are again left to the readers. In general, we expect to revisit the existing algorithms to ensure that they are meaningful for genuine mathematical sequences or functions, and not only for algebraic idealizations.

Another big challenge in this project originates in the scientific difference between computer algebra and formal proofs. Computer algebra seeks speed of calculation on *concrete instances* of algebraic data structures (polynomials, matrices, etc). For their part, formal proofs manipulate symbolic expressions in terms of *abstract variables* understood to represent generic elements of algebraic data structures. In view of this, a continuous challenge is to develop the right, hybrid thinking attitude that is able to effectively manage concrete and abstract values simultaneously, alternatively computing and proving with them.

2.1.3. Experimental mathematics with special functions

Applications in combinatorics and mathematical physics frequently involve equations of so high orders and so large sizes, that computing or even storing all their coefficients is impossible on existing computers. Making this tractable is an extraordinary challenge. The approach we believe in is to design algorithms of good—ideally quasi-optimal—complexity in order to extract precisely the required data from the equations, while avoiding the computationally intractable task of completely expanding them into an explicit representation.

Typical applications with expected high impact are the automatic discovery and algorithmic proof of results in combinatorics and mathematical physics for which human proofs are currently unattainable.

2.2. Research axes

The implementation of certified symbolic computations on special functions in the Coq proof assistant requires both investigating new formalization techniques and renewing the traditional computer-algebra viewpoint on these standard objects. Large mathematical objects typical of computer algebra occur during formalization, which also requires us to improve the efficiency and ergonomics of Coq. In order to feed this interdisciplinary activity with new motivating problems, we additionally pursue a research activity oriented towards experimental mathematics in application domains that involve special functions. We expect these applications to pose new algorithmic challenges to computer algebra, which in turn will deserve a formal-certification effort. Finally, DDMF is the motivation and the showcase of our progress on the certification of these computations. While striving to provide a formal guarantee of the correctness of the information it displays, we remain keen on enriching its mathematical content by developing new computer-algebra algorithms.

2.2.1. Computer algebra certified by the Coq system

Our formalization effort consists in organizing a cooperation between a computer-algebra system and a proof assistant. The computer-algebra system is used to produce efficiently algebraic data, which are later processed by the proof assistant. The success of this cooperation relies on the design of appropriate libraries of formalized mathematics, including certified implementations of certain computer-algebra algorithms. On the other side, we expect that scrutinizing the implementation and the output of computer-algebra algorithms will shed a new light on their semantics and on their correctness proofs, and help clarifying their documentation.

2.2.1.1. Libraries of formalized mathematics

The appropriate framework for the study of efficient algorithms for special functions is *algebraic*. Representing algebraic theories as Coq formal libraries takes benefit from the methodology emerging from the success of ambitious projects like the formal proof of a major classification result in finite-group theory (the Odd Order Theorem) [40].

Yet, a number of the objects we need to formalize in the present context has never been investigated using any interactive proof assistant, despite being considered as commonplaces in computer algebra. For instance there is up to our knowledge no available formalization of the theory of non-commutative rings, of the algorithmic theory of special-functions closures, or of the asymptotic study of special functions. We expect our future formal libraries to prove broadly reusable in later formalizations of seemingly unrelated theories.

2.2.1.2. Manipulation of large algebraic data in a proof assistant

Another peculiarity of the mathematical objects we are going to manipulate with the Coq system is their size. In order to provide a formal guarantee on the data displayed by DDMF, two related axes of research have to be pursued. First, efficient algorithms dealing with these large objects have to be programmed and run in Coq. Recent evolutions of the Coq system to improve the efficiency of its internal computations [21], [24] make this objective reachable. Still, how to combine the aforementioned formalization methodology with these cutting-edge evolutions of Coq remains one of the prospective aspects of our project. A second need is to help users *interactively* manipulate large expressions occurring in their conjectures, an objective for which little has been done so far. To address this need, we work on improving the ergonomics of the system in two ways:

first, ameliorating the reactivity of Coq in its interaction with the user; second, designing and implementing extensions of its interface to ease our formalization activity. We expect the outcome of these lines of research to be useful to a wider audience, interested in manipulating large formulas on topics possibly unrelated to special functions.

2.2.1.3. *Formal-proof-producing normalization algorithms*

Our algorithm certifications inside Coq intend to simulate well-identified components of our Maple packages, possibly by reproducing them in Coq. It would however not have been judicious to re-implement them inside Coq in a systematic way. Indeed for a number of its components, the output of the algorithm is more easily checked than found, like for instance the solving of a linear system. Rather, we delegate the discovery of the solutions to an external, untrusted oracle like Maple. Trusted computations inside Coq then formally validate the correctness of the a priori untrusted output. More often than not, this validation consists in implementing and executing normalization procedures *inside* Coq. A challenge of this automation is to make sure they go to scale while remaining efficient, which requires a Coq version of non-trivial computer-algebra algorithms. A first, archetypal example we expect to work on is a non-commutative generalization of the normalization procedure for elements of rings [46].

2.2.2. Better symbolic computations with special functions

Generally speaking, we design algorithms for manipulating special functions symbolically, whether univariate or with parameters, and for extracting algorithmically any kind of algebraic and analytic information from them, notably asymptotic properties. Beyond this, the heart of our research is concerned with parametrised definite summations and integrations. These very expressive operations have far-ranging applications, for instance, to the computation of integral transforms (Laplace, Fourier) or to the solution of combinatorial problems expressed via integrals (coefficient extractions, diagonals). The algorithms that we design for them need to really operate on the level of linear functional systems, differential and of recurrence. In all cases, we strive to design our algorithms with the constant goal of good theoretical complexity, and we observe that our algorithms are also fast in practice.

2.2.2.1. *Special-function integration and summation*

Our long-term goal is to design fast algorithms for a general method for special-function integration (*creative telescoping*), and make them applicable to general special-function inputs. Still, our strategy is to proceed with simpler, more specific classes first (rational functions, then algebraic functions, hyperexponential functions, D-finite functions, non-D-finite functions; two variables, then many variables); as well, we isolate analytic questions by first considering types of integration with a more purely algebraic flavor (constant terms, algebraic residues, diagonals of combinatorics). In particular, we expect to extend our recent approach [29] to more general classes (algebraic with nested radicals, for example): the idea is to speed up calculations by making use of an analogue of Hermite reduction that avoids considering certificates. Homologous problems for summation will be addressed as well.

2.2.2.2. *Applications to experimental mathematics*

As a consequence of our complexity-driven approach to algorithms design, the algorithms mentioned in the previous paragraph are of good complexity. Therefore, they naturally help us deal with applications that involve equations of high orders and large sizes.

With regard to combinatorics, we expect to advance the algorithmic classification of combinatorial classes like walks and urns. Here, the goal is to determine if enumerative generating functions are rational, algebraic, or D-finite, for example. Physical problems whose modelling involves special-function integrals comprise the study of models of statistical mechanics, like the Ising model for ferro-magnetism, or questions related to Hamiltonian systems.

Number theory is another promising domain of applications. Here, we attempt an experimental approach to the automated certification of integrality of the coefficients of mirror maps for Calabi–Yau manifolds. This could also involve the discovery of new Calabi–Yau operators and the certification of the existing ones. We also plan to algorithmically discover and certify new recurrences yielding good approximants needed in irrationality proofs.

It is to be noted that in all of these application domains, we would so far use general algorithms, as was done in earlier works of ours [28], [32], [31]. To push the scale of applications further, we plan to consider in each case the specifics of the application domain to tailor our algorithms.

2.2.3. Interactive and certified mathematical web sites

In continuation of our past project of an encyclopedia at http://ddmf.msr-inria.inria.fr/1.9.1/ddmf, we ambition to both enrich and certify the formulas about the special functions that we provide online. For each function, our website shows its essential properties and the mathematical objects attached to it, which are often infinite in nature (numerical evaluations, asymptotic expansions). An interactive presentation has the advantage of allowing for adaption to the user's needs. More advanced content will broaden the encyclopedia:

- the algorithmic discussion of equations with parameters, leading to certified automatic case analysis based on arithmetic properties of the parameters;
- lists of summation and integral formulas involving special functions, including validity conditions on the parameters;
- guaranteed large-precision numerical evaluations.

3. Research Program

3.1. Studying special functions by computer algebra

Computer algebra manipulates symbolic representations of exact mathematical objects in a computer, in order to perform computations and operations like simplifying expressions and solving equations for "closed-form expressions". The manipulations are often fundamentally of algebraic nature, even when the ultimate goal is analytic. The issue of efficiency is a particular one in computer algebra, owing to the extreme swell of the intermediate values during calculations.

Our view on the domain is that research on the algorithmic manipulation of special functions is anchored between two paradigms:

- adopting linear differential equations as the right data structure for special functions,
- designing efficient algorithms in a complexity-driven way.

It aims at four kinds of algorithmic goals:

- algorithms combining functions,
- functional equations solving,
- multi-precision numerical evaluations,
- guessing heuristics.

This interacts with three domains of research:

- computer algebra, meant as the search for quasi-optimal algorithms for exact algebraic objects,
- symbolic analysis/algebraic analysis;
- experimental mathematics (combinatorics, mathematical physics, ...).

This view is made explicit in the present section.

3.1.1. Equations as a data structure

Numerous special functions satisfy linear differential and/or recurrence equations. Under a mild technical condition, the existence of such equations induces a finiteness property that makes the main properties of the functions decidable. We thus speak of *D-finite functions*. For example, 60 % of the chapters in the handbook [18] describe D-finite functions. In addition, the class is closed under a rich set of algebraic operations. This makes linear functional equations just the right data structure to encode and manipulate special functions. The power of this representation was observed in the early 1990s [70], leading to the design of many algorithms in computer algebra. Both on the theoretical and algorithmic sides, the study of D-finite functions shares much with neighbouring mathematical domains: differential algebra, D-module theory, differential Galois theory, as well as their counterparts for recurrence equations.

3.1.2. Algorithms combining functions

Differential/recurrence equations that define special functions can be recombined [70] to define: additions and products of special functions; compositions of special functions; integrals and sums involving special functions. Zeilberger's fast algorithm for obtaining recurrences satisfied by parametrised binomial sums was developed in the early 1990s already [71]. It is the basis of all modern definite summation and integration algorithms. The theory was made fully rigorous and algorithmic in later works, mostly by a group in RISC (Linz, Austria) and by members of the team [59], [67], [35], [33], [34], [54]. The past ÉPI Algorithms contributed several implementations (*gfun* [62], *Mgfun* [35]).

3.1.3. Solving functional equations

Encoding special functions as defining linear functional equations postpones some of the difficulty of the problems to a delayed solving of equations. But at the same time, solving (for special classes of functions) is a sub-task of many algorithms on special functions, especially so when solving in terms of polynomial or rational functions. A lot of work has been done in this direction in the 1990s; more intensively since the 2000s, solving differential and recurrence equations in terms of special functions has also been investigated.

3.1.4. Multi-precision numerical evaluation

A major conceptual and algorithmic difference exists for numerical calculations between data structures that fit on a machine word and data structures of arbitrary length, that is, *multi-precision* arithmetic. When multi-precision floating-point numbers became available, early works on the evaluation of special functions were just promising that "most" digits in the output were correct, and performed by heuristically increasing precision during intermediate calculations, without intended rigour. The original theory has evolved in a twofold way since the 1990s: by making computable all constants hidden in asymptotic approximations, it became possible to guarantee a *prescribed* absolute precision; by employing state-of-the-art algorithms on polynomials, matrices, etc, it became possible to have evaluation algorithms in a time complexity that is linear in the output size, with a constant that is not more than a few units. On the implementation side, several original works exist, one of which (*NumGfun* [58]) is used in our DDMF.

3.1.5. Guessing heuristics

"Differential approximation", or "Guessing", is an operation to get an ODE likely to be satisfied by a given approximate series expansion of an unknown function. This has been used at least since the 1970s and is a key stone in spectacular applications in experimental mathematics [32]. All this is based on subtle algorithms for Hermite–Padé approximants [22]. Moreover, guessing can at times be complemented by proven quantitative results that turn the heuristics into an algorithm [30]. This is a promising algorithmic approach that deserves more attention than it has received so far.

3.1.6. Complexity-driven design of algorithms

The main concern of computer algebra has long been to prove the feasibility of a given problem, that is, to show the existence of an algorithmic solution for it. However, with the advent of faster and faster computers, complexity results have ceased to be of theoretical interest only. Nowadays, a large track of works in computer algebra is interested in developing fast algorithms, with time complexity as close as possible to linear in their output size. After most of the more pervasive objects like integers, polynomials, and matrices have been endowed with fast algorithms for the main operations on them [41], the community, including ourselves, started to turn its attention to differential and recurrence objects in the 2000s. The subject is still not as developed as in the commutative case, and a major challenge remains to understand the combinatorics behind summation and integration. On the methodological side, several paradigms occur repeatedly in fast algorithms: "divide and conquer" to balance calculations, "evaluation and interpolation" to avoid intermediate swell of data, etc. [27].

3.2. Trusted computer-algebra calculations

3.2.1. Encyclopedias

Handbooks collecting mathematical properties aim at serving as reference, therefore trusted, documents. The decision of several authors or maintainers of such knowledge bases to move from paper books [18], [20], [63] to websites and wikis [0] allows for a more collaborative effort in proof reading. Another step toward further confidence is to manage to generate the content of an encyclopedia by computer-algebra programs, as is the case with the Wolfram Functions Site [0] or DDMF [0]. Yet, due to the lingering doubts about computer-algebra systems, some encyclopedias propose both cross-checking by different systems and handwritten companion paper proofs of their content [0]. As of today, there is no encyclopedia certified with formal proofs.

3.2.2. Computer algebra and symbolic logic

Several attempts have been made in order to extend existing computer-algebra systems with symbolic manipulations of logical formulas. Yet, these works are more about extending the expressivity of computer-algebra systems than about improving the standards of correctness and semantics of the systems. Conversely, several projects have addressed the communication of a proof system with a computer-algebra system, resulting in an increased automation available in the proof system, to the price of the uncertainty of the computations performed by this oracle.

3.2.3. Certifying systems for computer algebra

More ambitious projects have tried to design a new computer-algebra system providing an environment where the user could both program efficiently and elaborate formal and machine-checked proofs of correctness, by calling a general-purpose proof assistant like the Coq system. This approach requires a huge manpower and a daunting effort in order to re-implement a complete computer-algebra system, as well as the libraries of formal mathematics required by such formal proofs.

3.2.4. Semantics for computer algebra

The move to machine-checked proofs of the mathematical correctness of the output of computer-algebra implementations demands a prior clarification about the often implicit assumptions on which the presumably correctly implemented algorithms rely. Interestingly, this preliminary work, which could be considered as independent from a formal certification project, is seldom precise or even available in the literature.

3.2.5. Formal proofs for symbolic components of computer-algebra systems

A number of authors have investigated ways to organize the communication of a chosen computer-algebra system with a chosen proof assistant in order to certify specific components of the computer-algebra systems, experimenting various combinations of systems and various formats for mathematical exchanges. Another line of research consists in the implementation and certification of computer-algebra algorithms inside the logic [66], [46], [55] or as a proof-automation strategy. Normalization algorithms are of special interest when they allow to check results possibly obtained by an external computer-algebra oracle [38]. A discussion about the systematic separation of the search for a solution and the checking of the solution is already clearly outlined in [52].

3.2.6. Formal proofs for numerical components of computer-algebra systems

Significant progress has been made in the certification of numerical applications by formal proofs. Libraries formalizing and implementing floating-point arithmetic as well as large numbers and arbitrary-precision arithmetic are available. These libraries are used to certify floating-point programs, implementations of mathematical functions and for applications like hybrid systems.

---

[0] for instance http://dlmf.nist.gov/ for special functions or http://oeis.org/ for integer sequences
[0] http://functions.wolfram.com/
[0] http://ddmf.msr-inria.inria.fr/1.9.1/ddmf
[0] http://129.81.170.14/~vhm/Table.html

3.3. Machine-checked proofs of formalized mathematics

To be checked by a machine, a proof needs to be expressed in a constrained, relatively simple formal language. Proof assistants provide facilities to write proofs in such languages. But, as merely writing, even in a formal language, does not constitute a formal proof just per se, proof assistants also provide a proof checker: a small and well-understood piece of software in charge of verifying the correctness of arbitrarily large proofs. The gap between the low-level formal language a machine can check and the sophistication of an average page of mathematics is conspicuous and unavoidable. Proof assistants try to bridge this gap by offering facilities, like notations or automation, to support convenient formalization methodologies. Indeed, many aspects, from the logical foundation to the user interface, play an important role in the feasibility of formalized mathematics inside a proof assistant.

3.3.1. Logical foundations and proof assistants

While many logical foundations for mathematics have been proposed, studied, and implemented, type theory is the one that has been more successfully employed to formalize mathematics, to the notable exception of the Mizar system [56], which is based on set theory. In particular, the calculus of construction (CoC) [36] and its extension with inductive types (CIC) [37], have been studied for more than 20 years and been implemented by several independent tools (like Lego, Matita, and Agda). Its reference implementation, Coq [64], has been used for several large-scale formalizations projects (formal certification of a compiler back-end; four-color theorem). Improving the type theory underlying the Coq system remains an active area of research. Other systems based on different type theories do exist and, whilst being more oriented toward software verification, have been also used to verify results of mainstream mathematics (prime-number theorem; Kepler conjecture).

3.3.2. Computations in formal proofs

The most distinguishing feature of CoC is that computation is promoted to the status of rigorous logical argument. Moreover, in its extension CIC, we can recognize the key ingredients of a functional programming language like inductive types, pattern matching, and recursive functions. Indeed, one can program effectively inside tools based on CIC like Coq. This possibility has paved the way to many effective formalization techniques that were essential to the most impressive formalizations made in CIC.

Another milestone in the promotion of the computations-as-proofs feature of Coq has been the integration of compilation techniques in the system to speed up evaluation. Coq can now run realistic programs in the logic, and hence easily incorporates calculations into proofs that demand heavy computational steps.

Because of their different choice for the underlying logic, other proof assistants have to simulate computations outside the formal system, and indeed fewer attempts to formalize mathematical proofs involving heavy calculations have been made in these tools. The only notable exception, which was finished in 2014, the Kepler conjecture, required a significant work to optimize the rewriting engine that simulates evaluation in Isabelle/HOL.

3.3.3. Large-scale computations for proofs inside the Coq system

Programs run and proved correct inside the logic are especially useful for the conception of automated decision procedures. To this end, inductive types are used as an internal language for the description of mathematical objects by their syntax, thus enabling programs to reason and compute by case analysis and recursion on symbolic expressions.

The output of complex and optimized programs external to the proof assistant can also be stamped with a formal proof of correctness when their result is easier to *check* than to *find*. In that case one can benefit from their efficiency without compromising the level of confidence on their output at the price of writing and certify a checker inside the logic. This approach, which has been successfully used in various contexts, is very relevant to the present research project.

3.3.4. Relevant contributions from the Mathematical Component libraries

Representing abstract algebra in a proof assistant has been studied for long. The libraries developed by the MathComp project for the proof of the Odd Order Theorem provide a rather comprehensive hierarchy of structures; however, they originally feature a large number of instances of structures that they need to organize. On the methodological side, this hierarchy is an incarnation of an original work [40] based on various mechanisms, primarily type inference, typically employed in the area of programming languages. A large amount of information that is implicit in handwritten proofs, and that must become explicit at formalization time, can be systematically recovered following this methodology.

Small-scale reflection [43] is another methodology promoted by the MathComp project. Its ultimate goal is to ease formal proofs by systematically dealing with as many bureaucratic steps as possible, by automated computation. For instance, as opposed to the style advocated by Coq's standard library, decidable predicates are systematically represented using computable boolean functions: comparison on integers is expressed as program, and to state that $a \leq b$ one compares the output of this program run on a and b with $true$. In many cases, for example when a and b are values, one can prove or disprove the inequality by pure computation.

The MathComp library was consistently designed after uniform principles of software engineering. These principles range from simple ones, like naming conventions, to more advanced ones, like generic programming, resulting in a robust and reusable collection of formal mathematical components. This large body of formalized mathematics covers a broad panel of algebraic theories, including of course advanced topics of finite group theory, but also linear algebra, commutative algebra, Galois theory, and representation theory. We refer the interested reader to the online documentation of these libraries [65], which represent about 150,000 lines of code and include roughly 4,000 definitions and 13,000 theorems.

Topics not addressed by these libraries and that might be relevant to the present project include real analysis and differential equations. The most advanced work of formalization on these domains is available in the HOL-Light system [48], [49], [50], although some existing developments of interest [25], [57] are also available for Coq. Another aspect of the MathComp libraries that needs improvement, owing to the size of the data we manipulate, is the connection with efficient data structures and implementations, which only starts to be explored.

3.3.5. User interaction with the proof assistant

The user of a proof assistant describes the proof he wants to formalize in the system using a textual language. Depending on the peculiarities of the formal system and the applicative domain, different proof languages have been developed. Some proof assistants promote the use of a declarative language, when the Coq and Matita systems are more oriented toward a procedural style.

The development of the large, consistent body of MathComp libraries has prompted the need to design an alternative and coherent language extension for the Coq proof assistant [45], [44], enforcing the robustness of proof scripts to the numerous changes induced by code refactoring and enhancing the support for the methodology of small-scale reflection.

The development of large libraries is quite a novelty for the Coq system. In particular any long-term development process requires the iteration of many refactoring steps and very little support is provided by most proof assistants, with the notable exception of Mizar [61]. For the Coq system, this is an active area of research.

4. Application Domains

4.1. Computer Algebra in Mathematics

Our expertise in computer algebra and complexity-driven design of algebraic algorithms has applications in various domains, including:

- combinatorics, especially the study of combinatorial walks,
- theoretical computer science, like by the study of automatic sequences,
- number theory, by the analysis of the nature of so-called periods.

5. New Software and Platforms

5.1. DynaMoW

Dynamic Mathematics on the Web

FUNCTIONAL DESCRIPTION: Programming tool for controlling the generation of mathematical websites that embed dynamical mathematical contents generated by computer-algebra calculations. Implemented in OCaml.

- Participants: Alexis Darrasse, Frédéric Chyzak and Maxence Guesdon
- Contact: Frédéric Chyzak
- URL: http://ddmf.msr-inria.inria.fr/DynaMoW/

5.2. ECS

Encyclopedia of Combinatorial Structures

FUNCTIONAL DESCRIPTION: On-line mathematical encyclopedia with an emphasis on sequences that arise in the context of decomposable combinatorial structures, with the possibility to search by the first terms in the sequence, keyword, generating function, or closed form.

- Participants: Alexis Darrasse, Frédéric Chyzak, Maxence Guesdon and Stéphanie Petit
- Contact: Frédéric Chyzak
- URL: http://ecs.inria.fr/

5.3. DDMF

Dynamic Dictionary of Mathematical Functions

FUNCTIONAL DESCRIPTION: Web site consisting of interactive tables of mathematical formulas on elementary and special functions. The formulas are automatically generated by OCaml and computer-algebra routines. Users can ask for more terms of the expansions, more digits of the numerical values, proofs of some of the formulas, etc.

- Participants: Alexandre Benoit, Alexis Darrasse, Bruno Salvy, Christoph Koutschan, Frédéric Chyzak, Marc Mezzarobba, Maxence Guesdon, Stefan Gerhold and Thomas Gregoire
- Contact: Frédéric Chyzak
- URL: http://ddmf.msr-inria.inria.fr/1.9.1/ddmf

5.4. Mgfun

multivariate generating functions package

FUNCTIONAL DESCRIPTION: The Mgfun Project is a collection of packages for the computer algebra system Maple, and is intended for the symbolic manipulation of a large class of special functions and combinatorial sequences (in one or several variables and indices) that appear in many branches of mathematics, mathematical physics, and engineering sciences. Members of the class satisfy a crucial finiteness property which makes the class amenable to computer algebra methods and enjoy numerous algorithmic closure properties, including algorithmic closures under integration and summation.

- Contact: Frédéric Chyzak
- URL: http://specfun.inria.fr/chyzak/mgfun.html

5.5. Ssreflect

KEYWORD: Proof assistant

SCIENTIFIC DESCRIPTION: Ssreflect is tactic language that helps writing concise and uniform tactic based proof scripts for the Coq system. It was designed during the proofs of the 4 Color Theorem and the Feit-Thompson theorem.

FUNCTIONAL DESCRIPTION: Ssreflect is a tactic language extension to the Coq system, developed by the Mathematical Components team.

NEWS OF THE YEAR: In 2019, we extended the intro pattern functionality of SSreflect and added support for working under binders using the "under" tactical.

- Participants: Assia Mahboubi, Cyril Cohen, Enrico Tassi, Georges Gonthier, Laurence Rideau, Laurent Théry and Yves Bertot
- Contact: Yves Bertot
- URL: http://math-comp.github.io/math-comp/

5.6. Math-Components

Mathematical Components library

KEYWORD: Proof assistant

FUNCTIONAL DESCRIPTION: The Mathematical Components library is a set of Coq libraries that cover the prerequiste for the mechanization of the proof of the Odd Order Theorem.

RELEASE FUNCTIONAL DESCRIPTION: This releases is compatible with Coq 8.9 and Coq 8.10 it adds many theorems for finite function, prime numbers, sequences, finite types, bigo operations, natural numbers, cycles in graphs.

- Participants: Alexey Solovyev, Andrea Asperti, Assia Mahboubi, Cyril Cohen, Enrico Tassi, François Garillot, Georges Gonthier, Ioana Pasca, Jeremy Avigad, Laurence Rideau, Laurent Théry, Russell O'Connor, Sidi Ould Biha, Stéphane Le Roux and Yves Bertot
- Contact: Assia Mahboubi
- URL: http://math-comp.github.io/math-comp/

6. New Results

6.1. Becker's conjecture on Mahler functions

In 1994, Becker conjectured that if $F(z)$ is a k-regular power series, then there exists a k-regular rational function $R(z)$ such that $F(z)/R(z)$ satisfies a Mahler-type functional equation with polynomial coefficients, whose trailing coefficient (i.e., of order 0) is 1. In [2], Frédéric Chyzak and Philippe Dumas, together with Jason P. Bell (University of Waterloo, Canada) and Michael Coons (University of Newcastle, Australia) have proved Becker's conjecture in the best-possible form: they have shown that the rational function $R(z)$ can be taken to be a polynomial $z^\gamma Q(z)$ for some explicit non-negative integer γ and such that $1/Q(z)$ is k-regular. The article was published this year.

6.2. Fast coefficient computation for algebraic power series in positive characteristic

In [8], Alin Bostan and Philippe Dumas, together with Xavier Caruso (CNRS, Rennes) and Gilles Christol (IMJ, Paris) have studied the algorithmic question of coefficient computation of algebraic power series in positive characteristic. They revisited Christol's theorem on algebraic power series in positive characteristic and proposed another proof for it. Their new proof combines several ingredients and advantages of existing proofs, which make it very well-suited for algorithmic purposes. The construction used in the new proof was then applied to the design of a new efficient algorithm for computing the Nth coefficient of a given algebraic power series over a perfect field of characteristic p. This algorithm has several nice features: it is more general, more natural and more efficient than previous algorithms. Not only the arithmetic complexity of the new algorithm is linear in $\log N$ and quasi-linear in p, but its dependency with respect to the degree of the input is much smaller than in the previously best algorithm. Moreover, when the ground field is finite, the new approach yields an even faster algorithm, whose bit complexity is linear in $\log N$ and quasi-linear in $\sqrt{p}$.

6.3. Subresultants of $(x-\alpha)^m$ and $(x-\beta)^n$, Jacobi polynomials and complexity

A previous article described explicit expressions for the coefficients of the order-d polynomial subresultant of $(x-\alpha)^m$ and $(x-\beta)^n$ with respect to Bernstein's set of polynomials $\{(x-\alpha)^j(x-\beta)^{d-j},\ 0\le j\le d\}$, for $0\le d<\min\{m,n\}$. In [3], Alin Bostan, together with T. Krick, M. Valdettaro (U. Buenos Aires, Argentina) and A. Szanto (U. North Carolina, Raleigh, USA) further developed the study of these structured polynomials and showed that the coefficients of the subresultants of $(x-\alpha)^m$ and $(x-\beta)^n$ with respect to the monomial basis can be computed in *linear* arithmetic complexity, which is faster than for arbitrary polynomials. The result is obtained as a consequence of the amazing though seemingly unnoticed fact that these subresultants are scalar multiples of Jacobi polynomials up to an affine change of variables.

6.4. Least common multiple of random integers

In [4], Alin Bostan together with Kilian Raschel (CNRS, Tours) and Alexander Marynych (U. Kyiv, Ukraine) have investigated the least common multiple of random integers. Using a purely probabilistic approach, they derived a criterion for the convergence in distribution as $n\to\infty$ of $f(L_n)/n^{rk}$, for a wide class of multiplicative arithmetic functions f with polynomial growth r, where $L_n(k)$ denotes the least common multiple of k independent random integers with uniform distribution on $\{1,2,...,n\}$. Furthermore, they identified the limit as an infinite product of independent random variables indexed by the prime numbers. Along the way of showing the main results, they computed the (rational) generating function of a trimmed sum of independent geometric laws, which appears in the above infinite product. The latter is directly related to the generating function of a certain max-type diophantine equation, of which they solved a generalized version. The results extend theorems by Erdős and Wintner (1939), Fernández and Fernández (2013) and Hilberdink and Tóth (2016).

6.5. On sequences associated to the invariant theory of rank two simple Lie algebras

In [14], Alin Bostan together with Jordan Tirrell (Washington College, USA) Philadelphia, USA), Bruce W. Westbury (Unversity of Texas at Dallas, USA) and Yi Zhang (Xi'an Jiaotong-Liverpool University, Suzhou, China) studied two families of sequences, listed in the On-Line Encyclopedia of Integer Sequences (OEIS), which are associated to invariant theory of Lie algebras. For the first family, they proved combinatorially that the sequences A059710 and A108307 are related by a binomial transform. Based on this, they presented two independent proofs of a recurrence equation for A059710, which was conjectured by Mihailovs. Besides, they also gave a direct proof of Mihailovs' conjecture by the method of algebraic residues. As a consequence, closed formulae for the generating function of sequence A059710 were obtained in terms of classical Gaussian hypergeometric functions.

6.6. Explicit degree bounds for right factors of linear differential operators

If a linear differential operator with rational function coefficients is reducible, its factors may have coefficients with numerators and denominators of very high degree. When the base field is $\mathbb{C}$, Alin Bostan together with Bruno Salvy (Inria and ENS Lyon) and Tanguy Rivoal (CNRS and U. Grenoble) gave in [13] a completely explicit bound for the degrees of the monic right factors in terms of the degree and the order of the original operator, as well as the largest modulus of the local exponents at all its singularities. As a consequence, if a differential operator L has rational function coefficients over a number field, they obtained degree bounds for its monic right factors in terms of the degree, the order and the height of L, and of the degree of the number field.

6.7. Improved algorithms for left factorial residues

In [11], Alin Bostan together with Vladica Andrejić (University of Belgrade, Serbia) and Milos Tatarevic (CoinList, Alameda, CA) presented improved algorithms for computing the left factorial residues $!p = 0! + 1! + \cdots + (p-1)! \bmod p$. They used these algorithms for the calculation of the residues $!p \bmod p$, for all primes p up to 2^{40}. Their results confirm that Kurepa's left factorial conjecture is still an open problem, as they show that there are no odd primes $p < 2^{40}$ such that p divides $!p$. Additionally, they confirmed that there are no socialist primes p with $5 < p < 2^{40}$.

6.8. A note on gamma triangles and local gamma vectors

Alin Bostan contributed to F. Chapoton's article [5] by writing an appendix, which allowed the author to complete its article. The theme of [5] is the study of simplicial complexes in algebraic combinatorics. A basic invariant is the f-vector that counts faces according to their dimensions. A less understood invariant is the γ-vector, introduced by Gal in 2005. Also in 2005, Chapoton, motivated by the study of the combinatorics of simplicial complexes attached to cluster algebras, considered a refined version of the f-vector. The main aim of [5] is to introduce the analogue in this context of the γ-vector, and a further refinement called the Γ-triangle. The author computed explicitly the Γ-triangle for all the cluster simplicial complexes of irreducible Coxeter groups. Alin Bostan contributed to the proof of an unexpected relation between the Γ-triangles of cluster fans of type $\mathbb{B}$ and $\mathbb{D}$.

6.9. A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions

In the preliminary work [16], Frédéric Chyzak and Frank Nielsen (LIX, Palaiseau and Sony Computer Science Laboratories, Tokyo, Japan) have reported on a closed-form expression for the Kullback-Leibler divergence between Cauchy distributions which involves the calculation of a parametric definite integral with 6 parameters. The formula shows that the Kullback-Leibler divergence between Cauchy densities is always finite and symmetric. This work also serves as a show-case of several methods in computer algebra to the computation of parametrized integrals.

6.10. Big prime field FFT on multi-core processors

In [9], Svyatoslav Covanov, together with Davood Mohajerani, Marc Moreno Maza, and Linxiao Wang (all from ORCCA, Canada), have worked on a multi-threaded implementation of Fast Fourier Transforms over generalized Fermat prime fields. This work extends their previous study realized on graphics processing units to multi-core processors. In this new context, they overcome the less fine control of hardware resources by successively using FFT in support of the multiplication in those fields. They obtain favorable speedup factors (up to 6.9$\times$ on a 6-core, 12 threads node, and 4.3$\times$ on a 4-core, 8 threads node) of their parallel implementation compared to the serial implementation for the overall application thanks to the low memory footprint and the sharp control of arithmetic instructions of their implementation of generalized Fermat prime fields.

6.11. Martin boundary of killed random walks on isoradial graphs

Alin Bostan contributed to an article by Cédric Boutillier and Kilian Raschel [15], devoted to the study of random walks on isoradial graphs. Contrary to the lattice case, isoradial graphs are not translation invariant, do not admit any group structure and are spatially non-homogeneous. However, Boutillier and Raschel have been able to obtain analogues of a celebrated result by Ney and Spitzer (1966) on the so-called *Martin kernel* (ratio of Green functions started at different points). Alin Bostan provided in the Appendix two different proofs of the fact that some algebraic power series arising in this context have non-negative coefficients.

6.12. Random walks in orthants and lattice path combinatorics

In the second edition of the book [39], original methods were proposed to determine the invariant measure of random walks in the quarter plane with small jumps (size 1), the general solution being obtained via reduction to boundary value problems. Among other things, an important quantity, the so-called *group of the walk*, allows to deduce theoretical features about the nature of the solutions. In particular, when the order of the group is finite and the underlying algebraic curve is of genus 0 or 1, necessary and sufficient conditions have been given for the solution to be rational, algebraic or D-finite (i.e., solution of a linear differential equation). In this framework, a number of difficult open problems related to lattice-path combinatorics are currently being explored by Alin Bostan, Frédéric Chyzak, and Guy Fayolle, both from the theoretical and computer-algebra viewpoints: concrete computation of the criteria, utilization of differential Galois theory, genus greater than 1 (i.e., when some jumps are of size ≥ 2), etc. A recent topic of future research deals with the connections between simple product-form stochastic networks (so-called *Jackson networks*) and explicit solutions of functional equations for counting lattice walks, see [17].

6.13. Quasilinear Average Complexity for Solving Polynomial Systems

How many operations do we need on the average to compute an approximate root of a random Gaussian polynomial system? Beyond Smale's 17th problem that asked whether a polynomial bound is possible, Pierre Lairez has proved in [6] a quasi-optimal bound $(inputsize)^{1+o(1)}$, which improves upon the previously known $(inputsize)^{3/2+o(1)}$ bound. His new algorithm relies on numerical continuation along *rigid continuation paths*. The central idea is to consider rigid motions of the equations rather than line segments in the linear space of all polynomial systems. This leads to a better average condition number and allows for bigger steps. He showed that on the average, one approximate root of a random Gaussian polynomial system of n equations of degree at most D in $n+1$ homogeneous variables can be computed with $O(n^5 D^2)$ continuation steps. This is a decisive improvement over previous bounds, which prove no better than $\sqrt{2}^{\min(n,D)}$ continuation steps on the average.

In 2019, the article has been accepted in the Journal of the AMS.

6.14. Computing the Volume of Compact Semi-Algebraic Sets

In [10], Pierre Lairez, Mohab Safey El Din and Marc Mezzarobba join a unique set of expertise in symbolic integration, real algebraic geometry and numerical integration to tackle a problem as old as Babylonian mathematics: the computation of volumes.

Let $S \subset R^n$ be a compact basic semi-algebraic set defined as the real solution set of multivariate polynomial inequalities with rational coefficients. They design an algorithm which takes as input a polynomial system defining S and an integer $p \geq 0$ and returns the n-dimensional volume of S at absolute precision 2^{-p}.

Their algorithm relies on the relationship between volumes of semi-algebraic sets and periods of rational integrals. It makes use of algorithms computing the Picard-Fuchs differential equation of appropriate periods, properties of critical points, and high-precision numerical integration of differential equations.

The algorithm runs in essentially linear time with respect to p. This improves upon the previous exponential bounds obtained by Monte-Carlo or moment-based methods.

6.15. Densities of Stieltjes moment sequences for pattern-avoiding permutations

A small subset of combinatorial sequences have coefficients that can be represented as moments of a nonnegative measure on $[0, \infty)$. Such sequences are known as *Stieltjes moment sequences*. They have a number of useful properties, such as log-convexity, which in turn enables one to rigorously bound their growth constant from below.

In [12], Alin Bostan together with Andrew Elvey Price, Anthony Guttmann and Jean-Marie Maillard, studied some classical sequences in enumerative combinatorics, denoted $Av(\mathcal{P})$, and counting permutations of $\{1, 2, ..., n\}$ that avoid some given pattern $\mathcal{P}$. For increasing patterns $\mathcal{P} = (12...k)$, they showed that the corresponding sequences, $Av(123...k)$, are Stieltjes moment sequences, and explicitly determined the underlying density function, either exactly or numerically, by using the Stieltjes inversion formula as a fundamental tool.

They showed that the densities for $Av(1234)$ and $Av(12345)$, correspond to an order-one linear differential operator acting on a classical modular form given as a pullback of a Gaussian ${}_2F_1$ hypergeometric function, respectively to an order-two linear differential operator acting on the square of a classical modular form given as a pullback of a ${}_2F_1$ hypergeometric function. Moreover, these density functions are closely, but non-trivially, related to the density attached to the distance traveled by a walk in the plane with $k - 1$ unit steps in random directions.

As a bonus, they studied the challenging case of the $Av(1324)$ sequence and gave compelling numerical evidence that this too is a Stieltjes moment sequence. Accepting this, they proved new lower bounds on the growth constant of this sequence, which are stronger than existing bounds. A further unproven assumption leads to even better bounds, which can be extrapolated to give a good estimate of the (unknown) growth constant.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR

- *De rerum natura*. This project, set up by the team, was accepted this year and will be funded until 2023. It gathers over 20 experts from four fields: computer algebra; the Galois theories of linear functional equations; number theory; combinatorics and probability. Our goal is to obtain classification algorithms for number theory and combinatorics, particularly so for deciding irrationality and transcendence.

7.1.2. Research in Pairs

Alin Bostan together with Marc Mezzaroba (CNRS, Sorbonne Université) and Tanguy Rivoal (CNRS, Université Grenoble-Alpes) have done a “research in pairs” on the Fast Computation of Values of D-Finite Functions, from December 2 to 6, 2019, at CIRM (Luminy, France). The aim of the joint project was to investigate the implications of arithmetic properties of linear differential equations on the computational complexity of their numerical solutions. They focussed on E- and G-functions, which are power series solutions of differential equations that additionally satisfy strong arithmetic conditions and play a major role in Diophantine approximation. The main goal for this research session was to understand several remarks, given without proof by Chudnovsky and Chudnovsky in the late 1980s, and stating that number-theoretic properties could lead to slightly better complexity bounds for E- and G-functions than in the general case.

7.2. International Research Visitors

7.2.1. Visits of International Scientists

7.2.1.1. Internships

- Pierre Lairez supervised during two months Abhijit Balachandra, M1-level student from the Indian Institute of Science (Bangalore). They studied some new aspects of the numerical computation of the topology of complex algebraic surfaces.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

8.1.1.1. General Chair, Scientific Chair

- Alin Bostan is part of the Scientific advisory board of the conference series *Effective Methods in Algebraic Geometry* (MEGA).
- Alin Bostan is part of the scientific committee of the GDR EFI ("Functional Equations and Interactions") dependent on the mathematical institute (INSMI) of the CNRS. The goal of this GDR is to bring together various research communities in France working on functional equations in fields of computer science and mathematics.
- Frédéric Chyzak is member of the steering committee of the *Journées Nationales de Calcul Formel* (JNCF), the annual meeting of the French computer algebra community.
- Frédéric Chyzak was until July 2019 elected member (and chair) of the steering committee of the *International Symposium on Symbolic and Algebraic Computation* (ISSAC, 3-year term).
- Georges Gonthier is a member of the steering committee of the *Certified Programs and Proofs* Conference (CPP).

8.1.1.2. Member of the Organizing Committees

- Alin Bostan co-organizes, with Lucia Di Vizio, the *Séminaire Différentiel* between U. Versailles and Inria Saclay, with a bi-annual frequency ($\sim$ 30 participants per event).
- Alin Bostan co-organizes, with Lucia Di Vizio, the working group *Marches dans le quart de plan*, at Institut Henri Poincaré (Paris), with a bi-monthly frequency ($\sim$ 15 participants per event).

8.1.2. Scientific Events: Selection

8.1.2.1. Member of the Conference Program Committees

- Alin Bostan and Frédéric Chyzak have served as conference program committee members for the first *Maple Conference*.
- Georges Gonthier has served as a conference program committee members for the first *Workshop on Formal Methods for Blockchains* (FMBC).

8.1.2.2. Reviewer

- Frédéric Chyzak has served as reviewer for the selection of the international conferences CICM 2019, ISSAC 2019, and Maple Conference 2019.
- Alin Bostan has served as reviewer for the selection of the international conferences FPSAC 2019 and Maple Conference 2019.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

- Alin Bostan is on the editorial board of the *Journal of Symbolic Computation*.
- Alin Bostan is on the editorial board of the *Annals of Combinatorics*.
- Guy Fayolle is associate editor of the journal *Markov Processes and Related Fields*.
- Georges Gonthier is on the editorial board of the *Journal of Formalized Reasoning*.

8.1.3.2. Reviewer - Reviewing Activities

- Alin Bostan has served as a reviewer for the journals: *Journal of Symbolic Computation*, *Journal of Combinatorial Theory, Series A*, *Applicable Algebra in Engineering Communications and Computing*, *Journal of Combinatorial Algebra*, *Annales Henri Lebesgue*, *Annali dell Università di Ferrara*, *Mathematics of Computation*, *Séminaire Lotharingien de Combinatoire*.
- Guy Fayolle has been a reviewer for *Advances in Applied Probability*, *Markov Processes and Related Fields*, *Probability Theory and Related Fields*, *Queueing Systems: Theory and Applications*, *European Journal of Combinatorics*, *Journal of Statistical Physics*, *Physica A*, *Springer Science*.

8.1.4. Invited Talks

- Frédéric Chyzak has been invited to give a talk on his joint work with Alin Bostan about the enumeration of walks with small steps in the quarter plane at the international conference Transient Transcendence in Transylvania (Braov, Romania).
- Frédéric Chyzak has been invited to give talks on his joint work with Philippe Dumas about Becker's conjecture on Mahler functions: at the conference Équations Fonctionnelles et Interactions (Anglet), during a Seminar on Symbolic Computation at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences (Beijing, China), and at the conference Differential Galois Theory in Strasbourg (Strasbourg).
- Frédéric Chyzak has been invited to give a talk on his joint work with Alin Bostan, Pierre Lairez, and Bruno Salvy (AriC) at the 6th Summer School in Symbolic Computation (Chongqing, China).
- Alin Bostan has been invited to give a talk at the *Algebraic Marvels in Differential Equations*, Universidade Lisboa, Lisbonne, Portugal, February 2019.
- Alin Bostan has been invited to give a talk at the Combinatorics Seminar, LaBRI, Bordeaux, March 2019.
- Alin Bostan has been a plenary speaker at the international conference AofA 2019, Luminy (France), June 2019.
- Alin Bostan has been a plenary speaker at the international conference FPSAC 2019, Ljubljana (Slovenia), July 2019.
- Alin Bostan has been invited to give a series of five lectures at the Vienna Summer School of Mathematics, Weissensee, Austria, Sept. 2019.

8.1.5. Leadership within the Scientific Community

8.1.5.1. Regular Research Seminar

The team organizes a regular seminar, with roughly 10 talks a year. The topics reflect the team's interests: computer algebra, combinatorics, number theory, formal proofs, and related domains.

8.1.5.2. Research Working Group

In 2018, we have set up a working group *Marches dans le quart de plan* around the study of walks in the quarter plan, a very active research topic in probability theory and enumerative combinatorics in recent years. The working group is organized at Institut Henri Poincaré, with a regularity of two sessions per month. The original purpose was to read the article "On the Nature of the Generating Series of Walks in the Quarter Plane" by T. Dreyfus, C. Hardouin, J. Roques, M. Singer, published in Invent. Math. this year. But the reality exceeded expectations: the working group attracted a dozen of people, working either in computer science or pure mathematics, who began to interact and a very good dynamic was created. Altogether, 15 sessions have taken place in 2019.

8.1.5.3. International Conference

Together with Kilian Raschel (CNRS, U. Tours), Alin Bostan co-organized an international conference, Transient Transcendence in Transylvania, held in Romania from May 13 to 17, 2019. They took care together of all the infrastructure for this conference: program, invitations, web page, etc. This conference was a unique event in Romania, with a truly exceptional list of speakers, from several continents and countries: South Africa, Germany, Austria, Canada, United States, France, Netherlands, Poland, and of course, Romania. As a natural continuation of the conference, a volume will be published in the Springer collection PROMS (Proceedings in Mathematics & Statistics), with Bostan and Raschel as editors.

8.1.6. Scientific Expertise

- Guy Fayolle is scientific advisor and associate researcher at the *Robotics Laboratory of Mines ParisTech*.
- Georges Gonthier is taking part in an interministerial survey on the technological roadblocks for blockchains, which has been jointly commissioned to Inria, CEA and IMT by the Ministère de l'Economie, the Ministère de l'Education supérieure et de la Recherche, and the Secrétariat d'Etat au Numérique. He also participates to the Blockchain Taskforce set up by the French government.

8.1.7. Research Administration

- Frédéric Chyzak is project coordinator of the ANR project *De rerum natura*.
- Guy Fayolle is a member of the working group for *Computer System Modeling* of the *International Federation for Information Processing* (IFIP WG 7.3).
- Georges Gonthier serves on the Conseil de l'École Doctorale de Mathématiques Hadamard.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master:

Alin Bostan, *Algorithmes efficaces en calcul formel*, 36h, M2, MPRI, France.

Alin Bostan, *Modern Algorithms for Symbolic Summation and Integration*, 21h, M2, Master d'Informatique Fondamentale de l'ENS de Lyon, France.

Frédéric Chyzak, *Algorithmes efficaces en calcul formel*, 22.5h, M2, MPRI, France.

Pierre Lairez, *Algorithmique avancée (INF550)*, TD, 18h, M2, École polytechnique, France.

Pierre Lairez, *Les bases de la programmation et de l'algorithmique (INF411)*, TD, 40h, M1, École polytechnique, France.

8.2.2. Juries

- Alin Bostan has served as an examiner in the PhD jury of Robin Larrieu, *Arithmétique rapide pour des corps finis*, Ecole polytechnique, December 10, 2019.
- Alin Bostan has served as a member of the monitoring PhD committee of Youssef Abdelaziz, Univ. Paris 6.
- Alin Bostan has served as a member of the monitoring PhD committee of Manon Bertin, Univ. Rouen.
- Frédéric Chyzak has served as a reviewer in the PhD jury of Joelle Saade, *Méthodes symboliques pour les systèmes différentiels linéaires à singularité irrégulière*, Université de Limoges, November 5, 2019.
- Frédéric Chyzak has served as a reviewer in the PhD jury of Amélie Trotignon, *Marches sur des réseaux dans des cônes : aspects combinatoires et probabilistes*, Université de Tours, December 6, 2019.

- Pierre Lairez has served as a reviewer in the PhD jury of Josué Tonelli-Cueto, *Condition and Homology in Semialgebraic Geometry*, TU Berlin, November 28, 2019.
- Georges Gonthier has served in the PhD jury of Armaël Guénaud, *Mechanized Verification of the Correctness and Asymptotic Complexity of Programs*, Université de Paris, December 16, 2019.

8.3. Popularization

8.3.1. Articles and contents

- Georges Gonthier published an interview article *Blockchain: ce que c'est, comment ça marche* in La Recherche, **545**, March 2019.
- Georges Gonthier co-wrote with Ivan Odonnat (Banque de France) *L'avenir du bitcoin et de la blockchain*, in Les Carnets de l'Institut Diderot (2019).

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] M. BARKATOU, S. MADDAH.*Formal solutions of singularly-perturbed linear differential systems*, in "Journal of Symbolic Computation", September 2019, vol. 94, p. 183-209 [*DOI :* 10.1016/J.JSC.2018.08.003], https://hal.archives-ouvertes.fr/hal-02393887

[2] J. P. BELL, F. CHYZAK, M. COONS, P. DUMAS.*Becker's conjecture on Mahler functions*, in "Transactions of the American Mathematical Society", 2019, vol. 372, p. 3405–3423, In press, forthcoming [*DOI :* 10.1090/TRAN/7762], https://hal.inria.fr/hal-01885598

[3] A. BOSTAN, T. KRICK, A. SZANTO, M. VALDETTARO.*Subresultants of $(x-\alpha)^m$ and $(x-\beta)^n$, Jacobi polynomials and complexity*, in "Journal of Symbolic Computation", 2019, forthcoming [*DOI :* 10.1016/J.JSC.2019.10.003], https://hal.archives-ouvertes.fr/hal-01966640

[4] A. BOSTAN, A. MARYNYCH, K. RASCHEL.*On the least common multiple of several random integers*, in "Journal of Number Theory", November 2019, vol. 204, p. 113–133, https://arxiv.org/abs/1901.03002 [*DOI :* 10.1016/J.JNT.2019.03.017], https://hal.archives-ouvertes.fr/hal-01984389

[5] F. CHAPOTON, A. BOSTAN.*A note on gamma triangles and local gamma vectors*, in "Annales de la Faculté des Sciences de Toulouse. Mathématiques.", 2019, https://arxiv.org/abs/1809.00575 , forthcoming, https://hal.archives-ouvertes.fr/hal-01866199

[6] P. LAIREZ.*Rigid continuation paths I. Quasilinear average complexity for solving polynomial systems*, in "Journal of the American Mathematical Society", 2019, forthcoming, https://hal.inria.fr/hal-01631778

[7] A. MAHBOUBI, G. MELQUIOND, T. SIBUT-PINOTE.*Formally Verified Approximations of Definite Integrals*, in "Journal of Automated Reasoning", February 2019, vol. 62, n° 2, p. 281-300 [*DOI :* 10.1007/S10817-018-9463-7], https://hal.inria.fr/hal-01630143

International Conferences with Proceedings

[8] A. BOSTAN, X. CARUSO, G. CHRISTOL, P. DUMAS.*Fast Coefficient Computation for Algebraic Power Series in Positive Characteristic*, in "ANTS-XIII - Thirteenth Algorithmic Number Theory Symposium", Madison, United States, R. SCHEIDLER, J. SORENSON (editors), Proceedings of the Thirteenth Algorithmic Number Theory Symposium (ANTS–XIII), Mathematical Sciences Publishers, 2019, vol. 2, n^{o} 1, p. 119-135, https://arxiv.org/abs/1806.06543 [*DOI :* 10.2140/OBS.2019.2-1], https://hal.archives-ouvertes.fr/hal-01816375

[9] S. COVANOV, D. MOHAJERANI, M. MORENO MAZA, L. WANG.*Big Prime Field FFT on Multi-core Processors*, in "ISSAC 2019 - International Symposium on Symbolic and Algebraic Computation", Pékin, China, July 2019, https://hal.inria.fr/hal-02191652

[10] P. LAIREZ, M. MEZZAROBBA, M. SAFEY EL DIN.*Computing the volume of compact semi-algebraic sets*, in "ISSAC 2019 - International Symposium on Symbolic and Algebraic Computation", Beijing, China, ACM, July 2019, https://arxiv.org/abs/1904.11705 , https://hal.archives-ouvertes.fr/hal-02110556

Other Publications

[11] V. ANDREJIĆ, A. BOSTAN, M. TATAREVIC.*Improved algorithms for left factorial residues*, December 2019, https://arxiv.org/abs/1904.09196 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02411741

[12] A. BOSTAN, A. ELVEY-PRICE, A. J. GUTTMANN, J.-M. MAILLARD.*Stieltjes moment sequences for pattern-avoiding permutations*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02425917

[13] A. BOSTAN, T. RIVOAL, B. SALVY.*Explicit degree bounds for right factors of linear differential operators*, July 2019, https://arxiv.org/abs/1906.05529 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02154679

[14] A. BOSTAN, J. TIRRELL, B. W. WESTBURY, Y. ZHANG.*On sequences associated to the invariant theory of rank two simple Lie algebras*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02411755

[15] C. BOUTILLIER, K. RASCHEL, A. BOSTAN.*Martin boundary of killed random walks on isoradial graphs*, December 2019, Avec un appendice d'Alin Bostan, https://hal.archives-ouvertes.fr/hal-02422417

[16] F. CHYZAK, F. NIELSEN.*A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions*, December 2019, https://arxiv.org/abs/1905.10965 - 8 pages, https://hal.inria.fr/hal-02420591

[17] G. FAYOLLE.*A note on the connection between product-form Jackson networks and counting lattice walks in the quarter plane*, January 2020, working paper or preprint, https://hal.inria.fr/hal-02415746

References in notes

[18] M. ABRAMOWITZ, I. A. STEGUN (editors). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover, New York, 1992, xiv+1046, Reprint of the 1972 edition

[19] *Computer Algebra Errors*, Article in mathematics blog MathOverflow, http://mathoverflow.net/questions/11517/computer-algebra-errors

[20] F. W. J. OLVER, D. W. LOZIER, R. F. BOISVERT, C. W. CLARK (editors). *NIST Handbook of mathematical functions*, Cambridge University Press, 2010

[21] M. ARMAND, B. GRÉGOIRE, A. SPIWACK, L. THÉRY.*Extending Coq with Imperative Features and its Application to SAT Verication*, in "Interactive Theorem Proving, international Conference, ITP 2010, Edinburgh, Scotland, July 11–14, 2010, Proceedings", Lecture Notes in Computer Science, Springer, 2010

[22] B. BECKERMANN, G. LABAHN.*A uniform approach for the fast computation of matrix-type Padé approximants*, in "SIAM J. Matrix Anal. Appl.", 1994, vol. 15, n^{o} 3, p. 804–823

[23] A. BENOIT, F. CHYZAK, A. DARRASSE, S. GERHOLD, M. MEZZAROBBA, B. SALVY.*The Dynamic Dictionary of Mathematical Functions (DDMF)*, in "The Third International Congress on Mathematical Software (ICMS 2010)", K. FUKUDA, J. VAN DER HOEVEN, M. JOSWIG, N. TAKAYAMA (editors), Lecture Notes in Computer Science, 2010, vol. 6327, p. 35–41, http://dx.doi.org/10.1007/978-3-642-15582-6_7

[24] M. BOESPFLUG, M. DÉNÈS, B. GRÉGOIRE.*Full reduction at full throttle*, in "First International Conference on Certified Programs and Proofs, Taiwan, December 7–9", Lecture Notes in Computer Science, Springer, 2011

[25] S. BOLDO, C. LELAY, G. MELQUIOND.*Improving Real Analysis in Coq: A User-Friendly Approach to Integrals and Derivatives*, in "Certified Programs and Proofs", C. HAWBLITZEL, D. MILLER (editors), Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, vol. 7679, p. 289-304, http://dx.doi.org/10.1007/978-3-642-35308-6_22

[26] S. BOLDO, G. MELQUIOND.*Flocq: A Unified Library for Proving Floating-point Algorithms in Coq*, in "Proceedings of the 20th IEEE Symposium on Computer Arithmetic", Tübingen, Germany, July 2011, p. 243–252

[27] A. BOSTAN.*Algorithmes rapides pour les polynômes, séries formelles et matrices*, in "Actes des Journées Nationales de Calcul Formel", Luminy, France, 2010, p. 75–262, Les cours du CIRM, tome 1, numéro 2, http://ccirm.cedram.org:80/ccirm-bin/fitem?id=CCIRM_2010__1_2_75_0

[28] A. BOSTAN, S. BOUKRAA, S. HASSANI, J.-M. MAILLARD, J.-A. WEIL, N. ZENINE.*Globally nilpotent differential operators and the square Ising model*, in "J. Phys. A: Math. Theor.", 2009, vol. 42, n^{o} 12, 50, http://dx.doi.org/10.1088/1751-8113/42/12/125206

[29] A. BOSTAN, S. CHEN, F. CHYZAK, Z. LI.*Complexity of creative telescoping for bivariate rational functions*, in "ISSAC'10: Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation", New York, NY, USA, ACM, 2010, p. 203–210, http://doi.acm.org/10.1145/1837934.1837975

[30] A. BOSTAN, F. CHYZAK, G. LECERF, B. SALVY, É. SCHOST.*Differential equations for algebraic functions*, in "ISSAC'07: Proceedings of the 2007 international symposium on Symbolic and algebraic computation", C. W. BROWN (editor), ACM Press, 2007, p. 25–32, http://dx.doi.org/10.1145/1277548.1277553

[31] A. BOSTAN, F. CHYZAK, M. VAN HOEIJ, L. PECH.*Explicit formula for the generating series of diagonal 3D rook paths*, in "Sém. Loth. Comb.", 2011, vol. B66a, 27, http://www.emis.de/journals/SLC/wpapers/s66bochhope.html

[32] A. BOSTAN, M. KAUERS.*The complete generating function for Gessel walks is algebraic*, in "Proceedings of the American Mathematical Society", September 2010, vol. 138, n^{o} 9, p. 3063–3078, With an appendix by Mark van Hoeij

[33] F. CHYZAK.*An extension of Zeilberger's fast algorithm to general holonomic functions*, in "Discrete Math.", 2000, vol. 217, n^{o} 1-3, p. 115–134, Formal power series and algebraic combinatorics (Vienna, 1997)

[34] F. CHYZAK, M. KAUERS, B. SALVY.*A Non-Holonomic Systems Approach to Special Function Identities*, in "ISSAC'09: Proceedings of the Twenty-Second International Symposium on Symbolic and Algebraic Computation", J. MAY (editor), 2009, p. 111–118, http://dx.doi.org/10.1145/1576702.1576720

[35] F. CHYZAK, B. SALVY.*Non-commutative elimination in Ore algebras proves multivariate identities*, in "J. Symbolic Comput.", 1998, vol. 26, n^{o} 2, p. 187–227

[36] T. COQUAND, G. P. HUET.*The Calculus of Constructions*, in "Inf. Comput.", 1988, vol. 76, n^{o} 2/3, p. 95-120, http://dx.doi.org/10.1016/0890-5401(88)90005-3

[37] T. COQUAND, C. PAULIN-MOHRING.*Inductively defined types*, in "Proceedings of Colog'88", P. MARTIN-LÖF, G. MINTS (editors), Lecture Notes in Computer Science, Springer-Verlag, 1990, vol. 417

[38] D. DELAHAYE, M. MAYERO.*Dealing with algebraic expressions over a field in Coq using Maple*, in "J. Symbolic Comput.", 2005, vol. 39, n^{o} 5, p. 569–592, Special issue on the integration of automated reasoning and computer algebra systems, http://dx.doi.org/10.1016/j.jsc.2004.12.004

[39] G. FAYOLLE, R. IASNOGORODSKI, V. A. MALYSHEV. , S. ASMUSSEN, P. W. GLYNN, Y. LE JAN (editors)*Random Walks in the Quarter Plane: Algebraic Methods, Boundary Value Problems, Applications to Queueing Systems and Analytic Combinatorics*, Probability Theory and Stochastic Modelling, Springer International Publishing, February 2017, vol. 40, 255, The first edition was published in 1999 [*DOI :* 10.1007/978-3-319-50930-3], https://hal.inria.fr/hal-01651919

[40] F. GARILLOT, G. GONTHIER, A. MAHBOUBI, L. RIDEAU.*Packaging Mathematical Structures*, in "Theorem Proving in Higher-Order Logics", S. BERGHOFER, T. NIPKOW, C. URBAN, M. WENZEL (editors), Lecture Notes in Computer Science, Springer, 2009, vol. 5674, p. 327–342

[41] J. VON ZUR. GATHEN, J. GERHARD.*Modern computer algebra*, 2nd, Cambridge University Press, New York, 2003, xiv+785

[42] G. GONTHIER.*Formal proofs—the four-colour theorem*, in "Notices of the AMS", 2008, vol. 55, n^{o} 11, p. 1382-1393

[43] G. GONTHIER, A. MAHBOUBI.*An introduction to small scale reflection in Coq*, in "Journal of Formalized Reasoning", 2010, vol. 3, n^{o} 2, p. 95–152

[44] G. GONTHIER, A. MAHBOUBI, E. TASSI.*A Small Scale Reflection Extension for the Coq system*, Inria, 2008, n^{o} RR-6455, http://hal.inria.fr/inria-00258384

[45] G. GONTHIER, E. TASSI.*A language of patterns for subterm selection*, in "ITP", LNCS, 2012, vol. 7406, p. 361–376

[46] B. GRÉGOIRE, A. MAHBOUBI.*Proving Equalities in a Commutative Ring Done Right in Coq*, in "Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005, Oxford, UK, August 22-25, 2005, Proceedings", Lecture Notes in Computer Science, Springer, 2005, vol. 3603, p. 98–113

[47] T. HALES.*Formal proof*, in "Notices of the AMS", 2008, vol. 55, n^{o} 11, p. 1370-1380

[48] J. HARRISON.*A HOL Theory of Euclidean space*, in "Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005", Oxford, UK, J. HURD, T. MELHAM (editors), Lecture Notes in Computer Science, Springer-Verlag, 2005, vol. 3603

[49] J. HARRISON.*Formalizing an analytic proof of the prime number theorem*, in "Journal of Automated Reasoning", 2009, vol. 43, p. 243–261, Dedicated to Mike Gordon on the occasion of his 60th birthday

[50] J. HARRISON.*Theorem proving with the real numbers*, CPHC/BCS distinguished dissertations, Springer, 1998

[51] J. HARRISON.*A Machine-Checked Theory of Floating Point Arithmetic*, in "Theorem Proving in Higher Order Logics: 12th International Conference, TPHOLs'99", Nice, France, Y. BERTOT, G. DOWEK, A. HIRSCHOWITZ, C. PAULIN, L. THÉRY (editors), Lecture Notes in Computer Science, Springer-Verlag, 1999, vol. 1690, p. 113–130

[52] J. HARRISON, L. THÉRY.*A Skeptic's Approach to Combining HOL and Maple*, in "J. Autom. Reason.", December 1998, vol. 21, n^{o} 3, p. 279–294, http://dx.doi.org/10.1023/A:1006023127567

[53] F. JOHANSSON.*Another Mathematica bug*, 2009, Article on personal blog, http://fredrik-j.blogspot.fr/2009/07/another-mathematica-bug.html

[54] C. KOUTSCHAN.*A fast approach to creative telescoping*, in "Math. Comput. Sci.", 2010, vol. 4, n^{o} 2-3, p. 259–266, http://dx.doi.org/10.1007/s11786-010-0055-0

[55] A. MAHBOUBI.*Implementing the cylindrical algebraic decomposition within the Coq system*, in "Mathematical Structures in Computer Science", 2007, vol. 17, n^{o} 1, p. 99–127

[56] R. MATUSZEWSKI, P. RUDNICKI.*Mizar: the first 30 years*, in "Mechanized Mathematics and Its Applications", 2005, vol. 4

[57] M. MAYERO.*Problèmes critiques et preuves formelles*, Université Paris 13, novembre 2012, Habilitation à Diriger des Recherches

[58] M. MEZZAROBBA.*NumGfun: a package for numerical and analytic computation and D-finite functions*, in "ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation", New York, ACM, 2010, p. 139–146, http://dx.doi.org/10.1145/1837934.1837965

[59] P. PAULE, M. SCHORN.*A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities*, in "J. Symbolic Comput.", 1995, vol. 20, n^{o} 5-6, p. 673–698, Symbolic computation in combinatorics Δ_1 (Ithaca, NY, 1993), http://dx.doi.org/10.1006/jsco.1995.1071

[60] B. PETERSEN.*Maple*, Personal web site

[61] P. RUDNICKI, A. TRYBULEC.*On the Integrity of a Repository of Formalized Mathematics*, in "Proceedings of the Second International Conference on Mathematical Knowledge Management", London, UK, MKM '03, Springer-Verlag, 2003, p. 162–174, http://dl.acm.org/citation.cfm?id=648071.748518

[62] B. SALVY, P. ZIMMERMANN.*Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable*, in "ACM Trans. Math. Software", 1994, vol. 20, n° 2, p. 163–177

[63] N. J. A. SLOANE, S. PLOUFFE.*The Encyclopedia of Integer Sequences*, Academic Press, San Diego, 1995

[64] THE COQ DEVELOPMENT TEAM.*The Coq Proof Assistant: Reference Manual*, http://coq.inria.fr/doc/

[65] THE MATHEMATICAL COMPONENT TEAM.*A Formalization of the Odd Order Theorem using the Coq proof assistant*, September 2012, http://www.msr-inria.fr/projects/mathematical-components/

[66] L. THÉRY.*A Machine-Checked Implementation of Buchberger's Algorithm*, in "J. Autom. Reasoning", 2001, vol. 26, n° 2, p. 107-137, http://dx.doi.org/10.1023/A:1026518331905

[67] K. WEGSCHAIDER.*Computer generated proofs of binomial multi-sum identities*, RISC, J. Kepler University, May 1997, 99

[68] S. WOLFRAM.*Mathematica: A system for doing mathematics by computer (2nd ed.)*, Addison-Wesley, 1992

[69] D. ZEILBERGER.*Opinion 94: The Human Obsession With "Formal Proofs" is a Waste of the Computer's Time, and, Even More Regretfully, of Humans' Time*, 2009, http://www.math.rutgers.edu/~zeilberg/Opinion94.html

[70] D. ZEILBERGER.*A holonomic systems approach to special functions identities*, in "J. Comput. Appl. Math.", 1990, vol. 32, n° 3, p. 321–368

[71] D. ZEILBERGER.*The method of creative telescoping*, in "J. Symbolic Comput.", 1991, vol. 11, n° 3, p. 195–204

Project-Team TAU

Tackling the under-specified

IN COLLABORATION WITH: Laboratoire de recherche en informatique (LRI)

IN PARTNERSHIP WITH:
CNRS

Université Paris-Sud (Paris 11)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team TAU

Creation of the Project-Team: 2019 July 01

Keywords:

Computer Science and Digital Science:
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.5.2. - Recommendation systems
A8.2. - Optimization
A8.6. - Information theory
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:
B1.1.4. - Genetics and genomics
B4. - Energy
B7.2.1. - Smart vehicles
B9.1.2. - Serious games
B9.5.3. - Physics
B9.5.5. - Mechanics
B9.5.6. - Data science
B9.6.10. - Digital humanities

1. Team, Visitors, External Collaborators

Research Scientists
Marc Schoenauer [Team leader, Inria, Senior Researcher, HDR]
Michele Alessandro Bucci [Inria, Starting Research Position, from Oct 2019]
Guillaume Charpiat [Inria, Researcher]
Cyril Furtlehner [Inria, Researcher]
Flora Jay [CNRS, Researcher]
Michèle Sebag [CNRS, Senior Researcher, HDR]
Paola Tubaro [CNRS, Researcher, HDR]

Faculty Members
Cécile Germain [Univ Paris-Sud, Emeritus, HDR]
Isabelle Guyon [Univ Paris-Sud and Inria, Professor and *Chaire d'Excellence UPSaclay-Inria*, HDR]
Nicolas Spyratos [Univ Paris-Sud, Emeritus, HDR]
Philippe Caillou [Univ Paris-Sud, Associate Professor]
Aurélien Decelle [Univ Paris-Saclay, Associate Professor]
Francois Landes [Univ Paris-Saclay, Associate Professor]

External Collaborators
Yann Ollivier [Facebook AI Research, Paris, HDR]
Olivier Goudet [Université d'Angers, HDR]
Véronique Ventos [Nukkai, Paris, HDR]
Wafa Medjitna [from Jun 2019 until Nov 2019]

Technical Staff

Victor Alfonso Naya [Inria, Engineer, from Dec 2019]
Berna Bakir Batu [Inria, Engineer]
Laurent Basara [Inria, Engineer]
Etienne Brame [Inria, Engineer, until Jan 2019]
Raphael Jaiswal [Inria, Engineer, until Nov 2019]
Adrien Pavao [Univ Paris-Sud, Engineer, from Nov 2019]
Jonas Renault [CNRS, Engineer]

PhD Students

Eleonore Bartenlian [Ministère de l'Enseignement Supérieur et de la Recherche]
Victor Berger [Inria]
Leonard Blier [Facebook]
Benjamin Donnot [RTE, until Jan 2019]
Mandar Chandorkar [CWI, until Oct 2019]
Balthazar Donon [RTE, from oct 2018]
Guillaume Doquet [Inria, until Dec 2019]
Victor Estrade [Univ Paris-Sud, until Sep 2019]
Loris Felardos Saint Jean [Inria]
Giancarlo Fissore [Univ Paris-Sud]
Julien Girard [CEA]
Diviyan Kalainathan [Univ Paris-Sud]
Armand Lacombe [Inria, from Oct 2019]
Wenzhuo Liu [IRT System X, from Nov 2019]
Zhengying Liu [Ecole Polytechnique]
Nizam Makdoud [Thales]
Marc Nabhan [Renault]
Adrian Pol [CERN]
Herilalaina Rakotoarison [Inria]
Théophile Sanchez [Univ Paris-Sud]
Lisheng Sun [Univ Paris-Sud]
Corentin Tallec [Ecole polytechnique, until Sep 2019]
Marion Ullmo [Univ Paris-Sud]
Pierre Wolinski [Ecole Normale Supérieure Paris]

Post-Doctoral Fellows

Jean Cury [Univ Paris-Sud]
Ksenia Gasnikova [Inria, from Feb 2019]
Saumya Jetley [Inria, from Oct 2019]

Visiting Scientist

Burak Yelmen [from Feb 2019 until Apr 2019]

Administrative Assistant

Laurence Fontana [Inria]

2. Overall Objectives

2.1. Presentation

Since its creation in 2003, TAO activities had constantly but slowly evolved, as old problems were being solved, and new applications arose, bringing new fundamental issues to tackle. But recent abrupt progresses in Machine Learning (and in particular in Deep Learning) have greatly accelerated these changes also within the team. It so happened that this change of slope also coincided with some more practical changes in TAO ecosystem: following Inria 12-years rule, the team definitely ended in December 2016. The new team TAU (for **TA**ckling the **U**nderspecified) has been proposed, and formally created in July 2019. At the same time important staff changes took place, that also justify even sharper changes in the team focus. During the year 2018, the second year of this new era for the (remaining) members of the team, our research topics have now stabilized around a final version of the TAU project.

Following the dramatic changes in TAU staff during the years 2016-2017 (see the 2017 activity report of the team for the details), the research around continuous optimization has definitely faded out in TAU (while the research axis on hyperparameter tuning has focused on Machine Learning algorithms), the Energy application domain has slightly changed direction under Isabelle Guyon's supervision (Section 4.2), after the completion of the work started by Olivier Teytaud, and a few new directions have emerged, around the robustness of ML systems (Section 3.1.2). The other research topics have been continued, as described below.

2.2. Context and overall goal of the project

Building upon the expertise in machine learning (ML) and optimization of the TAO team, the TAU project tackles some **under-specified challenges behind the New Artificial Intelligence wave**. The simultaneous advent of massive data and massive computational power, blurring the boundaries between data, structure, knowledge and common sense, seemingly makes it possible to fulfill all promises of the good old AI, now or soon.

This makes NewAI under-specified in three respects. A first dimension regards the relationships between AIs and human beings. The necessary conditions for AI sysytems to be accepted by mankind and/or contribute to the common good are yet to be formally defined; it is hard to believe that a general and computable definition of "ethical behavior" can be set once for all. Some of these necessary conditions (explainable and causal modeling; unbiased data and models; model certification) can nevertheless be cast as ambitious and realistic goals for public research.

A second dimension regards the relationships between AI, data and knowledge. Closed world AIs can manage and acquire sufficient data to reach human-level performances from scratch [81]. In open worlds however, prior knowledge is used in various ways to overcome the lack of direct interactions with the world, e.g. through i) exploiting domain-dependent data invariances in intension or in extension (ranging from convolution to domain augmentation); ii) taking advantage of the low-rank structure (generative learning) or known properties (equivariant learning) of the observed data; iii) leveraging diverse domains and datasets, assumedly related to each other (domain adaptation; multi-task learning). A general and open question is how available prior knowledge can be best leveraged by an AI algorithm, all the more so as domains with small to medium-size data are considered.

A third dimension regards the intrinsic limitations of AI in terms of information theory. Long established theories, e.g. rooted in Occam's razor, currently hardly account for the practical leaps of deep learning, where the solution dimension outnumbers the input dimension. Beyond trials-and-errors, a long-term goal is to characterize the learning landscape w.r.t. order parameters to be defined, and *a priori* estimate the regions of problem instances where it is likely/possible/unlikely to learn accurate models.

The above under-specified AI issues define three core research pillars (Section 3), examining three interdependent aspects of AI:

I. The first pillar aims to answer the question of what 'good AI' means, and how to build it. More specifically, our goal is to advance the state of the art concerning robust learning (against adversarial attacks), causal modeling (aimed to support explanations and prescriptions), and unbiased models in the sense of prescribed neutrality constraints (including the assessment and repair of the data).

II. The second pillar tackles the "innate *vs* acquired" question: how to best combine available human knowledge, and agnostic machine learning. TAU will examine this question focusing on domains with spatial and temporal multi-scale structure, as pervasive in natural sciences (where domain knowledge is expressed using PDEs, or through powerful compact representations as in signal processing), taking advantage of the pluri-disciplinary expertise and scientific collaborations of the TAU members.

III. The third pillar aims to understand the learning landscape. In the short term, it tackles the so-called Auto-☆ issue of automatically selecting and configuring an algorithm portfolio for a problem instance. This issue governs the knowledge transfer from research labs to industry [95], [94], all the more so as massive computational resources are at stake. In the medium term, our goal is to integrate hyper-parameters and model structure in the learning criteria, using information theory and/or bilevel programming [96]. In the long-term, our goal is to establish a phase diagram of the learning landscape, through i) determining order parameters; ii) relating the different regions defined along these order parameters, to the quality of the optimal solution, and the probability of finding a good approximation thereof. These goals are aligned with the unique scientific expertise of TAU in statistical physics and in information theory, and benefit from our decade-long expertise in Auto-☆.
The above research pillars will take inspiration and be validated with three application domains (Section 4):

1.Energy management encompasses a variety of scientific problems related to research pillars I. (fair learning, privacy-compliant modelling, safety-related guarantees) and II. (spatio-temporal multi-scale modelling, distributional learning). It is also a strategic application for the planet, where TAU benefits from the TAO expertise and the long established relationships with Artelys (ILab Metis) and RTE.

2.Computational Social Sciences offer questions and methodological lessons about how to address these questions in a common decency spirit, along research pillar I. On-going studies at TAU include the learning and randomized assessment of prescriptive models for Human Resources (hiring and vocational studies; quality of life at work and economic performance) and nutrition habits (in relation with social networks and health), where i) learned models must be unbiased although data are undoubtedly biased; ii) prior knowledge must be accounted for and the interpretation of the learned models is mandatory; iii) causal modelling is key as models are deployed for *prescription* and self-fulfilling prophecies must be avoided at all costs.

3.Optimal data-driven design considers several physical/simulated phenomenons, ranging from high-energy physics to space weather, from population biology to medical imaging, from signal processing to certification of autonomous vehicle controllers, with: i) medium-size data; ii) extensive prior knowledge, notably concerning the symmetries and properties of the sought models; iii) computationally expensive simulators. All three characteristics are relevant to pillars II and III.

3. Research Program

3.1. Toward Good AI

As discussed by [141], the topic of ethical AI was non-existent until 2010, was laughed at in 2016, and became a hot topic in 2017 as the AI disruptivity with respect to the fabric of life (travel, education, entertainment, social networks, politics, to name a few) became unavoidable [138], together with its expected impacts on the nature and amount of jobs. As of now, it seems that the risk of a new AI Winter might arise from legal[0] and societal[0] issues. While privacy is now recognized as a civil right in Europe, it is feared that the GAFAM, BATX and others can already capture a sufficient fraction of human preferences and their dynamics to achieve their commercial and other goals, and build a Brave New Big Brother (BNBB, a system that is openly beneficial to many, covertly nudging, and possibly dictatorial).

The ambition of TAU is to mitigate the BNBB risk along several intricated dimensions, and build i) causal and explainable models; ii) fair data and models; iii) provably robust models.

3.1.1. Causal modeling and biases

Participants: Isabelle Guyon, Michèle Sebag, Philippe Caillou, Paola Tubaro
PhD: Diviyan Kalainathan

[0] For instance, the (fictitious) plea challenge proposed to law students in Oct. 2018 considered a chain reaction pileup occurred among autonomous and humanly operated vehicles on a highway.

[0] For instance related to information bubbles and nudge [100], [155].

Collaboration: Olivier Goudet (Université d'Angers), David Lopez-Paz (Facebook)

The extraction of causal models, a long goal of AI [139], [117], [140], became a strategic issue as the usage of learned models gradually shifted from *prediction* to *prescription* in the last years. This evolution, following Auguste Comte's vision of science (*Savoir pour prévoir, afin de pouvoir*) indeed reflects the exuberant optimism about AI: Knowledge enables Prediction; Prediction enables Control. However, although predictive models can be based on correlations, prescriptions can only be based on causal models[0].

Among the research applications concerned with causal modeling, predictive modeling or collaborative filtering at TAU are all projects described in section 4.1 (see also Section 3.4), studying the relationships between: i) the educational background of persons and the job openings (FUI project JobAgile and DataIA project Vadore); ii) the quality of life at work and the economic performance indicators of the enterprises (ISN Lidex project Amiqap) [119] ; iii) the nutritional items bought by households (at the level of granularity of the barcode) and their health status, as approximated from their body-mass-index (IRS UPSaclay Nutriperso); iv) the actual offer of restaurants and their scores on online rating systems. In these projects, a wealth of data is available (though hardly sufficient for applications ii), iii and iv))) and there is little doubt that these data reflect the imbalances and biases of the world as is, ranging from gender to racial to economical prejudices. Preventing the learned models from perpetuating such biases is essential to deliver an AI endowed with common decency.

In some cases, the bias is known; for instance, the cohorts in the Nutriperso study are more well-off than the average French population, and the Kantar database includes explicit weights to address this bias through importance sampling. In other cases, the bias is only guessed; for instance, the companies for which Secafi data are available hardly correspond to a uniform sample as these data have been gathered upon the request of the company trade union.

3.1.2. Robustness of Learned Models

Participants: Guillaume Charpiat, Marc Schoenauer, Michèle Sebag
PhDs: Julien Girard, Marc Nabhan, Nizham Makhoud
Collaboration: Zakarian Chihani (CEA); Hiba Hage, and Yves Tourbier (Renault); Jérôme Kodjabachian (Thalès THERESIS)

Due to their outstanding performances, deep neural networks and more generally machine learning-based decision making systems, referred to as MLs in the following, have been raising hopes in the recent years to achieve breakthroughs in critical systems, ranging from autonomous vehicles to defense. The main pitfall for such applications lies in the lack of guarantees for MLs robustness.

Specifically, MLs are used when the mainstream software design process does not apply, that is, when no formal specification of the target software behavior is available and/or when the system is embedded in an open unpredictable world. The extensive body of knowledge developed to deliver guarantees about mainstream software – ranging from formal verification, model checking and abstract interpretation to testing, simulation and monitoring – thus does not directly apply either. Another weakness of MLs regards their dependency to the amount and quality of the training data, as their performances are sensitive to slight perturbations of the data distribution. Such perturbations can occur naturally due to domain or concept drift (e.g. due to a change in light intensity or a scratch on a camera lens); they can also result from intentional malicious attacks, a.k.a adversarial examples [156].

These downsides, currently preventing the dissemination of MLs in safety-critical systems (SCS), call for a considerable amount of research, in order to understand when and to which extent an MLs can be certified to provide the desired level of guarantees.

[0]One can predict that it rains based on the presence of umbrellas in the street; but one cannot induce rainfall by going out with an umbrella. Likewise, the presence of books/tablets at home and the good scores of children at school are correlated; but offering books/tablets to all children might fail to improve their scores *per se*, if both good scores and books are explained by a so-called confounder variable, like the presence of adults versed in books/tablets at home.

Julien Girard's PhD (CEA scholarship), started in Oct. 2018, co-supervised by Guillaume Charpiat and Zakaria Chihani (CEA), is devoted to the extension of abstract interpretation to deep neural nets, and the formal characterization of the transition kernel from input to output space achieved by a DNN (robustness by design, coupled with formally assessing the coverage of the training set). This approach is tightly related to the inspection and opening of black-box models, aimed to characterize the patterns in the input instances responsible for a decision – another step toward explainability.

On the other hand, experimental validation of MLs, akin statistical testing, also faces three limitations: i) real-world examples are notoriously insufficient to ensure a good coverage in general; ii) for this reason, simulated examples are extensively used; but their use raises the *reality gap* issue [128] of the distance between real and simulated worlds; iii) independently, the real-world is naturally subject to domain shift (e.g. due to the technical improvement and/or aging of sensors). Our collaborations with Renault tackle such issues in the context of the autonomous vehicle (see Section 7.1.3).

3.2. Hybridizing numerical modeling and learning systems

Participants: Alessandro Bucci, Guillaume Charpiat, Cécile Germain, Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhD: Théophile Sanchez, Loris Felardos, Wenzhuo Liu
In sciences and engineering, human knowledge is commonly expressed in closed form, through equations or mechanistic models characterizing how a natural or social phenomenon, or a physical device, will behave/evolve depending on its environment and external stimuli, under some assumptions and up to some approximations. The field of numerical engineering, and the simulators based on such mechanistic models, are at the core of most approaches to understand and analyze the world, from solid mechanics to computational fluid dynamics, from chemistry to molecular biology, from astronomy to population dynamics, from epidemiology and information propagation in social networks to economy and finance.

Most generally, numerical engineering supports the simulation, and when appropriate the optimization and control[0] of the phenomenons under study, although several sources of discrepancy might adversely affect the results, ranging from the underlying assumptions and simplifying hypotheses in the models, to systematic experiment errors to statistical measurement errors (not to mention numerical issues). This knowledge and know-how are materialized in millions of lines of code, capitalizing the expertise of academic and industrial labs. These softwares have been steadily extended over decades, modeling new and more fine-grained effects through layered extensions, making them increasingly harder to maintain, extend and master. Another difficulty is that complex systems most often resort to hybrid (pluridisciplinary) models, as they involve many components interacting along several time and space scales, hampering their numerical simulation.

At the other extreme, machine learning offers the opportunity to model phenomenons from scratch, using any available data gathered through experiments or simulations. Recent successes of machine learning in computer vision, natural language processing and games to name a few, have demonstrated the power of such agnostic approaches and their efficiency in terms of prediction [123], inverse problem solving [170], and sequential decision making [162], [81], despite their lack of any "semantic" understanding of the universe. Even before these successes, Anderson's claim was that *the data deluge [might make] the scientific method obsolete* [70], as if a reasonable option might be to throw away the existing equational or software bodies of knowledge, and let Machine Learning rediscover all models from scratch. Such a claim is hampered among others by the fact that not all domains offer a wealth of data, as any academic involved in an industrial collaboration around data has discovered.

Another approach will be considered in TAU, investigating how existing mechanistic models and related simulators can be partnered with ML algorithms: i) to achieve the same goals with the same methods with a gain of accuracy or time; ii) to achieve new goals; iii) to achieve the same goals with new methods.

[0]Note that the causal nature of mechanistic models is established from prior knowledge and experimentations.

Toward more robust numerical engineering: In domains where satisfying mechanistic models and simulators are available, ML can contribute to improve their accuracy or usability. A first direction is to refine or extend the models and simulators to better fit the empirical evidence. The goal is to finely account for the different biases and uncertainties attached to the available knowledge and data, distinguishing the different types of *known unknowns*. Such *known unknowns* include the model hyper-parameters (coefficients), the systematic errors due to e.g., experiment imperfections, and the statistical errors due to e.g., measurement errors. A second approach is based on learning a surrogate model for the phenomenon under study that incorporate domain knowledge from the mechanistic model (or its simulation). See Section 7.5 for case studies.

A related direction, typically when considering black-box simulators, aims to learn a model of the error, or equivalently, a post-processor of the software. The discrepancy between simulated and empirical results, referred to as *reality gap* [128], can be tackled in terms of domain adaptation [74], [99]. Specifically, the source domain here corresponds to the simulated phenomenon, offering a wealth of inexpensive data, and the target domain corresponds to the actual phenomenon, with rare and expensive data; the goal is to devise accurate target models using the source data and models.

Extending numerical engineering: ML, using both experimental and numerical data, can also be used to tackle new goals, that are beyond the current state-of-the-art of standard approaches. Inverse problems are such goals, identifying the parameters or the initial conditions of phenomenons for which the model is not differentiable, or amenable to the adjoint state method.

A slightly different kind of inverse problem is that of recovering the ground truth when only noisy data is available. This problem can be formulated as a search for the simplest model explaining the data. The question then becomes to formulate and efficiently exploit such a simplicity criterion.

Another goal can be to model the distribution of given quantiles for some system: The challenge is to exploit available data to train a generative model, aimed at sampling the target quantiles.

Examples tackled in TAU are detailed in Section 7.5. Note that the "Cracking the Glass Problem", described in Section 7.2.3 is yet another instance of a similar problem.

Data-driven numerical engineering : Finally, ML can also be used to sidestep numerical engineering limitations in terms of scalability, or to build a simulator emulating the resolution of the (unknown) mechanistic model from data, or to revisit the formal background.

When the mechanistic model is known and sufficiently accurate, it can be used to train a deep network on an arbitrary set of (space,time) samples, resulting in a meshless numerical approximation of the model [151], supporting by construction *differentiable programming* [125].

When no mechanistic model is sufficiently efficient, the model must be identified from the data only. Genetic programming has been used to identify systems of ODEs [149], through the identification of invariant quantities from data, as well as for the direct identification of control commands of nonlinear complex systems, including some chaotic systems [88]. Another recent approach uses two deep neural networks, one for the state of the system, the other for the equation itself [142]. The critical issues for both approaches include the scalability, and the explainability of the resulting models. Such line of research will benefit from TAU unique mixed expertise in Genetic Programming and Deep Learning.

Finally, in the realm of signal processing (SP), the question is whether and how deep networks can be used to revisit mainstream feature extraction based on Fourier decomposition, wavelet and scattering transforms [76]. E. Bartenlian's PhD (started Oct. 2018), co-supervised by M. Sebag and F. Pascal (Centrale-Supélec), focusing on musical audio-to-score translation [150], inspects the effects of supervised training, taking advantage from the fact that convolution masks can be initialized and analyzed in terms of frequency.

3.3. Learning to learn

According to Ali Rahimi's test of times award speech at NIPS 17, the current ML algorithms *have become a form of alchemy*. Competitive testing and empirical breakthroughs gradually become mandatory for a contribution to be acknowledged; an increasing part of the community adopts trials and errors as main scientific methodology, and theory is lagging behind practice. This style of progress is typical of technological and engineering revolutions for some; others ask for consolidated and well-understood theoretical advances, saving the time wasted in trying to build upon hardly reproducible results.

Basically, while practical achievements have often passed the expectations, there exist caveats along three dimensions. Firstly, excellent performances do not imply that the model has captured what was to learn, as shown by the phenomenon of adversarial examples. Following Ian Goodfellow, some well-performing models might be compared to *Clever Hans*, the horse that was able to solve mathematical exercizes using non verbal cues from its teacher [116]; it is the purpose of Pillar I. to alleviate the *Clever Hans* trap (section 3.1).
Secondly, some major advances, e.g. related to the celebrated adversarial learning [105], [99], establish proofs of concept more than a sound methodology, where the reproducibility is limited due to i) the computational power required for training (often beyond reach of academic labs); ii) the numerical instabilities (witnessed as random seeds happen to be found in the codes); iii) the insufficiently documented experimental settings. What works, why and when is still a matter of speculation, although better understanding the limitations of the current state of the art is acknowledged to be a priority. After Ali Rahimi again, *simple experiments, simple theorems are the building blocks that help us understand more complicated systems*. Along this line, [135] propose toy examples to demonstrate and understand the defaults of convergence of gradient descent adversarial learning.
Thirdly, and most importantly, the reported achievements rely on carefully tuned learning architectures and hyper-parameters. The sensitivity of the results to the selection and calibration of algorithms has been identified since the end 80s as a key ML bottleneck, and the field of automatic algorithm selection and calibration, referred to as AutoML or Auto-☆ in the following, is at the ML forefront.

TAU aims to contribute to the ML evolution toward a more mature stage along three dimensions. In the short term, the research done in Auto-☆ will be pursued (section 3.3.1). In the medium term, an information theoretic perspective will be adopted to capture the data structure and to calibrate the learning algorithm *depending on the nature and amount of the available data* (section 3.3.2). In the longer term, our goal is to leverage the methodologies forged in statistical physics to understand and control the trajectories of complex learning systems (section 3.3.3).

3.3.1. Auto-*

Participants: Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhD: Guillaume Doquet, Zhengying Liu, Herilalaina Rakotoarison, Lisheng Sun
Collaboration: Olivier Bousquet, André Elisseeff (Google Zurich)

The so-called Auto-☆ task, concerned with selecting a (quasi) optimal algorithm and its hyper-parameters depending on the problem instance at hand, remained a key issue in ML for the last three decades [75], as well as in optimization at large [115], including combinatorial optimization and constraint satisfaction [122], [104] and continuous optimization [71]. This issue, tackled by several European projects along the decades, governs the knowledge transfer to industry, due to the shortage of data scientists. It becomes even more crucial as models are more complex and their training requires more computational resources. This has motivated several international challenges devoted to Auto-ML [113] (see also Section 3.4), including the AutoDL challenge series [129] launched in 2019[0] (see also Section 7.6).

[0]https://autodl.chalearn.org/neurips2019

Several approaches have been used to tackle Auto-☆ in the literature, and TAU has been particularly active in several of them. Meta-learning aims to build a surrogate performance model, estimating the performance of an algorithm configuration on *any* problem instance characterized from its meta-feature values [146], [104], [72], [71], [103]. Collaborative filtering, considering that a problem instance "likes better" an algorithm configuration yielding a better performance, learns to recommend good algorithms to problem instances [153], [137]. Bayesian optimization proceeds by alternatively building a surrogate model of algorithm performances on *the* problem instance at hand, and tackling it [95]. This last approach currently is the prominent one; as shown in [137], the meta-features developed for AutoML are hardly relevant, hampering both meta-learning and collaborative filtering. The design of better features is another long-term research direction, in which TAU has recently been [32], ans still is very active. more recent approach used in TAU [40] extends the Bayesian Optimization approach with a Multi-Armed Bandit algorithm to generate the full Machine Learning pipeline, competing with the famed AutoSKLearn [95] (see Section 7.2.1). These results are presented in Section 7.2.1.

3.3.2. Information theory: adjusting model complexity and data fitting

Participants: Guillaume Charpiat, Marc Schoenauer, Michèle Sebag
PhD: Corentin Tallec, Pierre Wolinski, Léonard Blier
Collaboration: Yann Ollivier (Facebook)

In the 60s, Kolmogorov and Solomonoff provided a well-grounded theory for building (probabilistic) models best explaining the available data [147], [108], that is, the shortest programs able to generate these data. Such programs can then be used to generate further data or to answer specific questions (interpreted as missing values in the data). Deep learning, from this viewpoint, efficiently explores a space of computation graphs, described from its hyperparameters (network structure) and parameters (weights). Network training amounts to optimizing these parameters, namely, navigating the space of computational graphs to find a network, as simple as possible, that explain the past observations well.

This vision is at the core of variational auto-encoders [121], directly optimizing a bound on the Kolmogorov complexity of the dataset. More generally variational methods provide quantitative criteria to identify superfluous elements (edges, units) in a neural network, that can potentially be used for structural optimization of the network (Leonard Blier's PhD, started Oct. 2018).

The same principles apply to unsupervised learning, aimed to find the maximum amount of structure hidden in the data, quantified using this information-theoretic criterion.

The known invariances in the data can be exploited to guide the model design (e.g. as translation invariance leads to convolutional structures, or LSTM is shown to enforce the invariance to time affine transformations of the data sequence [157]). Scattering transforms exploit similar principles [76]. A general theory of how to detect *unknown* invariances in the data, however, is currently lacking.

The view of information theory and Kolmogorov complexity suggests that key program operations (composition, recursivity, use of predefined routines) should intervene when searching for a good computation graph. One possible framework for exploring the space of computation graphs with such operations is that of Genetic Programming. It is interesting to see that evolutionary computation appeared in the last two years among the best candidates to explore the space of deep learning structures [145], [126]. Other approaches might proceed by combining simple models into more powerful ones, e.g. using "Context Tree Weighting" [166] or switch distributions [90]. Another option is to formulate neural architecture design as a reinforcement learning problem [73]; the value of the building blocks (predefined routines) might be defined using e.g., Monte-Carlo Tree Search. A key difficulty is the computational cost of retraining neural nets from scratch upon modifying their architecture; an option might be to use neutral initializations to support warm-restart.

3.3.3. Analyzing and Learning Complex Systems

Participants: Cyril Furtlehner, Aurélien Decelle, François Landes, Michèle Sebag
PhD: Giancarlo Fissore

Collaboration: Enrico Camporeale (CWI); Jacopo Rocchi (LPTMS Paris Sud), the Simons team: Rahul Chako (post-doc), Andrea Liu (UPenn), David Reichman (Columbia), Giulio Biroli (ENS), Olivier Dauchot (ESPCI), Hufei Han (Symantec).

Methods and criteria from statistical physics have been widely used in ML. In early days, the capacity of Hopfield networks (associative memories defined by the attractors of an energy function) was investigated by using the replica formalism [69]. Restricted Boltzmann machines likewise define a generative model built upon an energy function trained from the data. Along the same lines, Variational Auto-Encoders can be interpreted as systems relating the free energy of the distribution, the information about the data and the entropy (the degree of ignorance about the micro-states of the system) [165]. A key promise of the statistical physics perspective and the Bayesian view of deep learning is to harness the tremendous growth of the model size (billions of weights in recent machine translation netwowrks), and make them sustainable through e.g. posterior drop-out [136], weight quantization and probabilistic binary networks [131]. Such "informational cooling" of a trained deep network can reduce its size by several orders of magnitude while preserving its performance.

Statistical physics is among the key expertises of TAU, originally only represented by Cyril Furtlehner, later strenghtened by Aurélien Decelle's and François Landes' arrivals in 2014 and 2018. On-going studies are conducted along several directions.

Generative models are most often expressed in terms of a Gibbs distributions $P[S] = \exp(-E[S])$, where energy E involves a sum of building blocks, modelling the interactions among variables. This formalization makes it natural to use mean-field methods of statistical physics and associated inference algorithms to both train and exploit such models. The difficulty is to find a good trade-off between the richness of the structure and the efficiency of mean-field approaches. One direction of research pursued in TAU, [97] in the context of traffic forecasting, is to account for the presence of cycles in the interaction graph, to adapt inference algorithms to such graphs with cycles, while constraining graphs to remain compatible with mean-field inference.

Another direction, explored in TAO/TAU in the recent years, is based on the definition and exploitation of self-consistency properties, enforcing principled divide-and-conquer resolutions. In the particular case of the message-passing Affinity Propagation algorithm for instance [168], self-consistency imposes the invariance of the solution when handled at different scales, thus enabling to characterize the critical value of the penalty and other hyper-parameters in closed form (in the case of simple data distributions) or empirically otherwise [98].

A more recent research direction examines the quantity of information in a (deep) neural net along the random matrix theory framework [78]. It is addressed in Giancarlo Fissore's PhD, and is detailed in Section 7.2.3.

Finally, we note the recent surge in using ML to address fundamental physics problems: from turbulence to high-energy physics and soft matter as well (with amorphous materials at its core) [19]. TAU's dual expertise in Deep Networks and in statistical physics places it in an ideal position to significantly contribute to this domain and shape the methods that will be used by the physics community in the future. François Landes' recent arrival in the team makes TAU a unique place for such interdisciplinary research, thanks to his collaborators from the Simons Collaboration *Cracking the Glass Problem* (gathering 13 statistical physics teams at the international level). This project is detailed in Section 7.2.3.

Independently, François Landes is actively collaborating with statistical physicists (Alberto Rosso, LPTMS, Univ. Paris-Saclay) and physcists at the frontier with geophysics (Eugenio Lippiello, Second Univ. of Naples) [20]. A possible CNRS grant (80Prime) may finance a shared PhD, at the frontier between seismicity and ML (Alberto Rosso, Marc Schoenauer and François Landes).

3.4. Organisation of Challenges

Participants: Cécile Germain, Isabelle Guyon, Marc Schoenauer, Michèle Sebag
Challenges have been an important drive for Machine Learning research for many years, and TAO members have played important roles in the organization of many such challenges: Michèle Sebag was head of the challenge programme in the Pascal European Network of Excellence (2005-2013); Isabelle Guyon, as mentioned, was the PI of many challenges ranging from causation challenges [109], to AutoML [110]. The Higgs challenge [67], most attended ever Kaggle challenge, was jointly organized by TAO (C. Germain),

LAL-IN2P3 (D. Rousseau and B. Kegl) and I. Guyon (not yet at TAO), in collaboration with CERN and Imperial College.

TAU was also particularly implicated with the ChaLearn Looking At People (LAP) challenge series in Computer Vision, in collaboration with the University of Barcelona [92] including the Job Candidate Screening Coopetition [91]; the Real Versus Fake Expressed Emotion Challenge (ICCV 2017) [163]; the Large-scale Continuous Gesture Recognition Challenge (ICCV 2017) [163]; the Large-scale Isolated Gesture Recognition Challenge (ICCV 2017) [163].

Other challenges have been organized in 2019, or are planned for the near future, detailed in Section 7.6. In particular, many of them now run on the Codalab platform, managed by Isabelle Guyon and maintained at LRI.

4. Application Domains

4.1. Computational Social Sciences

Participants: Philippe Caillou, Isabelle Guyon, Michèle Sebag, Paola Tubaro
Collaboration: Jean-Pierre Nadal (EHESS); Marco Cuturi, Bruno Crépon (ENSAE); Thierry Weil (Mines); Jean-Luc Bazet (RITM)

Computational Social Sciences (CSS) studies social and economic phenomena, ranging from technological innovation to politics, from media to social networks, from human resources to education, from inequalities to health. It combines perspectives from different scientific disciplines, building upon the tradition of computer simulation and modeling of complex social systems [102] on the one hand, and data science on the other hand, fueled by the capacity to collect and analyze massive amounts of digital data.

The emerging field of CSS raises formidable challenges along three dimensions. Firstly, the definition of the research questions, the formulation of hypotheses and the validation of the results require a tight pluridisciplinary interaction and dialogue between researchers from different backgrounds. Secondly, the development of CSS is a touchstone for ethical AI. On the one hand, CSS gains ground in major, data-rich private companies; on the other hand, public researchers around the world are engaging in an effort to use it for the benefit of society as a whole [124]. The key technical difficulties related to data and model biases, and to self-fulfilling prophecies have been discussed in section 3.1. Thirdly, CSS does not only regard scientists: it is essential that the civil society participate in the science of society [152].

TAO was involved in CSS for the last five years, and its activities have been strengthened thanks to P. Tubaro's and I. Guyon's expertises respectively in sociology and economics, and in causal modeling. Details are given in Section 7.3.

4.2. Energy Management

Participants: Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhD: Victor Berger, Benjamin Donnot, Balthazar Donon, Herilalaina Rakotoarison
Collaboration: Antoine Marot, Patrick Panciatici (RTE), Vincent Renault (Artelys)

Energy Management has been an application domain of choice for TAO since the end 2000s, with main partners SME Artelys (METIS Ilab Inria; ADEME project POST; on-going ADEME project NEXT), RTE (See.4C European challenge; two CIFRE PhDs), and, since Oct. 2019, IFPEN. The goals concern i) optimal planning over several spatio-temporal scales, from investments on continental Europe/North Africa grid at the decade scale (POST), to daily planning of local or regional power networks (NEXT); ii) monitoring and control of the French grid enforcing the prevention of power breaks (RTE); iii) improvement of house-made numerical methods using data-intense learning in all aspects of IFPEN activities (as described in Section 3.2).

Optimal planning over long periods of time amounts to optimal sequential decision under high uncertainties, ranging from stochastic uncertainties (weather, market prices, demand prediction) handled based on massive data, to non-stochastic uncertainties (e.g., political decisions about the nuclear policy) handled through defining and selecting a tractable number of scenarios. Note that non-anticipativity constraints forbid the use of dynamic programming-related methods; this led to propose the *Direct Value Search* method [77] at the end of the POST project.

The daily maintainance of power grids requires the building of approximate predictive models on the top of any given network topology. Deep Networks are natural candidates for such modelling, considering the size of the French grid ($\sim$ 10000 nodes), but the representation of the topology is a challenge when, e.g. the RTE goal is to quickly ensure the "n-1" security constraint (the network should remain safe even if any of the 10000 nodes fails). Existing simulators are too slow to be used in real time, and the size of actual grids makes it intractable to train surrogate models for all possible (n-1) topologies (see Section 7.4 for more details).

Furthermore, predictive models of local grids are based on the estimated consumption of end-customers: Linky meters only provide coarse grain information due to privacy issues, and very few samples of fine-grained consumption are available (from volunteer customers). A first task is to transfer knowledge from small data to the whole domain of application. A second task is to directly predict the peaks of consumption based on the user cluster profiles and their representativity (see Section 7.4.2).

4.3. Data-driven Numerical Modeling

Participants: Alessandro Bucci, Guillaume Charpiat, Cécile Germain, Isabelle Guyon, Flora Jay, Marc Schoenauer, Michèle Sebag
PhD and Post-doc: Victor Estrade, Loris Felardos, Adrian Pol, Théophile Sanchez, Wenzhuo Liu
Collaboration: D. Rousseau (LAL), M. Pierini (CERN)

As said (section 3.2), in domains where both first principle-based models and equations, and empirical or simulated data are available, their combined usage can support more accurate modelling and prediction, and when appropriate, optimization, control and design. This section describes such applications, with the goal of improving the time-to-design chain through fast interactions between the simulation, optimization, control and design stages. The expected advances regard: i) the quality of the models or simulators (through data assimilation, e.g. coupling first principles and data, or repairing/extending closed-form models); ii) the exploitation of data derived from different distributions and/or related phenomenons; and, most interestingly, iii) the task of optimal design and the assessment of the resulting designs.

The proposed approaches are based on generative and adversarial modelling [121], [106], extending both the generator and the discriminator modules to take advantage of the domain knowledge.

A first challenge regards the design of the model space, and the architecture used to enforce the known domain properties (symmetries, invariance operators, temporal structures). When appropriate, data from different distributions (e.g. simulated vs real-world data) will be reconciled, for instance taking inspiration from real-valued non-volume preserving transformations [85] in order to preserve the natural interpretation.
Another challenge regards the validation of the models and solutions of the optimal design problems. The more flexible the models, the more intensive the validation must be, as reminded by Leon Bottou. Along this way, generative models will be used to support the design of "what if" scenarios, to enhance anomaly detection and monitoring via refined likelihood criteria.

In the application case of dynamical systems such as fluid mechanics, the goal of incorporating machine learning into classical simulators is to speed up the simulations. Many possible tracks are possible for this; for instance one can search to provide better initialization heuristics to solvers (which make sure that physical constraints are satisfied, and which are responsible of most of the computational complexity of simulations) at each time step; one can also aim at predicting directly the state at $t + 100$, for instance, or at learning a representation space where the dynamics are linear (Koopman - von Neumann). The topic is very active in the deep learning community. To guarantee the quality of the predictions, concepts such as Liapunov coefficients (which express the speed at which simulated trajectories diverge from the true ones) can provide a suitable theoretical framework.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Best Paper Award in Machine Learning at ECML-PKDD 2019 in Wurzburg to Guillaume Doquet and Michèle Sebag for their paper *Agnostic feature selection* . See Guillaume's PhD [13] for more details.
- Nacim Belkhir, Winner ACM-GECCO 2019 **BBComp single-objective**, **BBComp two-objective** and **three-objective** tracks. The winning program is a slighly modified version of the one Nacim wrote during his PhD in TAU in 2017 [71], co-supervised by Marc Schoenauer, Johann Dréo and Pierre Savéant (Thalès TRT).

5.1.2. Visibility

- Marc Schoenauer, expert seconding Guillaume Klossa, special advisor to European Vice-President Ansip, for the report *Toward European Media Sovereignty* giving strategic advice on the opportunities and challenges related to the use artificial intelligence, with a focus on the media sector.

BEST PAPERS AWARDS :

[32]

G. F. DOQUET, M. SEBAG.*Agnostic feature selection*, in "ECML PKDD 2019 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Würzburg, Germany, September 2019, https://hal.archives-ouvertes.fr/hal-02436824

6. New Software and Platforms

6.1. io.datascience

Input Output Data Science

KEYWORDS: Open data - Semantic Web - FAIR (Findable, Accessible, Interoperable, and Reusable)

FUNCTIONAL DESCRIPTION: io.datascience (Input Output Data Science) is the instance of the Linked Wiki platform developed specifically in Paris-Saclay University as part of its Center for Data Science.

The goal of io.datascience: to facilitate the sharing and use of scientific data. The technological concept of io.datascience: the exploitation of semantic web advances, and in particular wiki technologies.

(Findable, Accessible, Interoperable, and Reusable) (Wilkinson, M., and The FAIR Guiding Principles for Scientific Data Management and Stewardship, Nature Scientific Data 2016)

io.datascience is both a data sharing platform and a framework for further development. It realizes a practical implementation of FAIR (Findable, Accessible, Interoperable, and Reusable - Wilkinson, M., Nature Scientific Data 2016) principles through a user-centric approach.

- Partners: Border Cloud - Paris Saclay Center for Data Science - Université Paris-Sud
- Contact: Cécile Germain
- Publications: Data acquisition for analytical platforms: Automating scientific workflows and building an open database platform for chemical anlysis metadata - A platform for scientific data sharing - TFT, Tests For Triplestores : Certifying the interoperability of RDF database systems using a continuous delivery workflow - Une autocomplétion générique de SPARQL dans un contexte multi-services - Certifying the interoperability of RDF database systems - Transforming Wikipedia into an Ontology-based Information Retrieval Search Engine for Local Experts using a Third-Party Taxonomy - The Grid Observatory 3.0 - Towards reproducible research and open collaborations using semantic technologies
- URL: https://io.datascience-paris-saclay.fr/

6.2. Codalab

KEYWORDS: Benchmarking - Competition

FUNCTIONAL DESCRIPTION: Challenges in machine learning and data science are competitions running over several weeks or months to resolve problems using provided datasets or simulated environments. Challenges can be thought of as crowdsourcing, benchmarking, and communication tools. They have been used for decades to test and compare competing solutions in machine learning in a fair and controlled way, to eliminate “inventor-evaluator" bias, and to stimulate the scientific community while promoting reproducible science. See our slide presentation.

As of June 2019 Codalab exceeded 40,000 users, 1000 competitions (300 public), and had over 300 submissions per day. Some of the areas in which Codalab is used include Computer vision and medical image analysis, natural language processing, time series prediction, causality, and automatic machine learning. Codalab was selected by the Région Ile de France to organize its challenges in the next three years.

TAU is going to continue expanding Codalab to accommodate new needs. One of our current focus is to support use of challenges for teaching (i.e. include a grading system as part of Codalab) and support for hooking up data simulation engines in the backend of Codalab to enable Reinforcement Learning challenges and simulate interactions of machines with an environment. For the fith year, we are using Codalab for student projects. M2 AIC students create mini data science challenges in teams of 6 students. L2 math and informatics students then solve them as part of their mini projects. We are collaborating with RPI (New York, USA) and Université de Grenoble to use this platform as part of a curriculum of medical students. We created a special application called ChaGrade to grade homework using challenges. Our PhD. students are involved in co-organizing challenges to expose the research community at large with the topic of their PhD. This helps them formalizing a task with rigor and allows them to disseminate their research.

- Contact: Isabelle Guyon
- URL: http://competitions.codalab.org

6.3. Cartolabe

KEYWORD: Information visualization

FUNCTIONAL DESCRIPTION: The goal of Cartolabe is to build a visual map representing the scientific activity of an institution/university/domain from published articles and reports. Using the HAL Database, Cartolabe provides the user with a map of the thematics, authors and articles . ML techniques are used for dimensionality reduction, cluster and topics identification, visualisation techniques are used for a scalable 2D representation of the results.

NEWS OF THE YEAR: This year, Cartolabe was applied to the Grand Debat dataset (3M individual propositions from french Citizen, see https://cartolabe.fr/map/debat). The results were used to test both the scaling capabilities of Cartolabe and its flexibility to non-scientific and non-english corpuses. We also Added submap capabilities to display the result of a year/lab/word filtering as an online generated heatmap with only the filtered points to facilitate the exploration.

- Participants: Philippe Caillou, Jean-Daniel Fekete, Jonas Renault and Anne-Catherine Letournel
- Partners: LRI - Laboratoire de Recherche en Informatique - CNRS
- Contact: Philippe Caillou
- URL: http://www.cartolabe.fr/

7. New Results

7.1. Toward Good AI

7.1.1. Causal Modeling

Participants: Philippe Caillou, Isabelle Guyon, Michèle Sebag;

PhDs: Diviyan Kalainathan
Collaboration: David Lopez-Paz (Facebook).

The search for **causal models** relies on quite a few hardly testable assumptions, e.g. causal sufficiency [160]; it is a data hungry task as it has the identification of independent and conditionally independent pairs of variables at its core. A new approach investigated through the Cause-Effects Pairs (CEP) Challenge [112] formulates causality search as a supervised learning problem, considering the joint distributions of pairs of variables (e.g. (Age, Salary)) labelled with the proper causation relationship between both variables (e.g. Age "causes" Salary) and learning algorithms apt to learn from distributions have been proposed [114]. An edited book has been published [48], that somewhat summarizes the whole history of Cause-EFfect Paris research. Several chapters of this book have co-authors in TAU: *Evaluation methods of cause-effect pairs* [49], *Learning Bivariate Functional Causal Models* [47], *Discriminant Learning Machines* [51], and *Results of the Cause-Effect Pair Challenge* [50].

In D. Kalainathan's PhD [14] and O. Goudet's postdoc, the search for causal models has been tackled in the framework of generative networks [107], trained to minimize the Maximum Mean Discrepancy loss; the resulting Causal Generative Neural Network improves on the state of the art on the CEP Challenge. CGNN favorably compares with the state of the art w.r.t. usual performance indicators (AUPR, SID) on main causal benchmarks, though with a large computational cost.

An attempt to scale up causal discovery, we proposed the Structural Agnostic Model approach [14] [120]. Working directly on the observational data, this global approach implements a variant of the popular adversarial game [99] between a discriminator, attempting to distinguish actual samples from fake ones, obtained by generating each variable, given real values from all others. A sparsity L_1 penalty forces all generators to consider only a small subset of their input variables, yielding a sparse causal graph. SAM obtains state-of-the-art performances on causal benchmarks, and scales up to a few hundred variables.

An innovative usage of causal models is for educational training in sensitive domains, such as medicine, along the following line. Given a causal generative model, artificial data can be generated using a marginal distribution of causes; such data will enable students to test their diagnosis inference (with no misleading spurious correlations in principle), while forbidding to reverse-engineer the artificial data and guess the original data. Some motivating applications for causal modeling are described in section 4.1.

7.1.2. Explainability

Participants: Isabelle Guyon, François Landes, Marc Schoenauer, Michèle Sebag
PhD: Marc Nabhan

Causal modeling is one particular method to tackle explainability, and TAU has been involved in other initiatives toward explainable AI systems. Following the LAP (Looking At People) challenges, Isabelle Guyon and co-organizers have edited a book [143] that presents a snapshot of explainable and interpretable models in the context of computer vision and machine learning. Along the same line, they propose an introduction and a complete survey of the state-of-the-art of the explainability and interpretability mechanisms in the context of first impressions analysis [57].

The team is also involved in the proposal for the IPL HyAIAI (Hybrid Approaches for Interpretable AI), coordinated by the LACODAM team (Rennes) dedicated to the design of hybrid approaches that combine state of the art numeric models (e.g., deep neural networks) with explainable symbolic models, in order to be able to integrate high level (domain) constraints in ML models, to give model designers information on ill-performing parts of the model, to provide understandable explanations on its results. Kickoff took place in September 2019, and we are still looking for good post-doc candidates.

Note also that the on-going work on the identification of the border of the failure zone in the parameter space of the autonomous vehicle simulator [37] (Section 7.1.3) also pertains to explainability.

Finally, a completely original approach to DNN explainability might arise from the study of structural glasses (7.2.3), with a parallel to Graph Neural Networks (GNNs), that could become an excellent non-trivial example for developing explainability protocols.

7.1.3. Robustness of AI Systems

Participants: Guillaume Charpiat, Marc Schoenauer, Michèle Sebag
PhDs: Julien Girard, Marc Nabhan, Nizham Makhoud
Collaboration: Zakarian Chihani (CEA); Hiba Hage and Yves Tourbier (Renault); Johanne Cohen (LRI-GALAC) and Christophe Labreuche (Thalès)

As said (Section 3.1.2, TAU is considering two directions of research related to the certification of MLs. The first direction, related to formal approaches, is the topic of Julien Girard's PhD (see also Section 3.1.2). On the opposite, the second axis aims to increase the robustness of systems that can only be experimentally validated. Two paths are investigated in the team: assessing the coverage of the datasets (more particularly here, used to train an autonomous vehicle controller), topic of Marc Nabhan's CIFRE with Renault; and detecting flaws in the system by reinforcement learning, as done by Nizam Makdoud's CIFRE PhD with Thalès THERESIS.

Formal validation of Neural Networks

The topic of provable deep neural network robustness has raised considerable interest in recent years. Most research in the literature has focused on adversarial robustness, which studies the robustness of perceptive models in the neighbourhood of particular samples. However, other works have proved global properties of smaller neural networks. Yet, formally verifying perception remains uncharted. This is due notably to the lack of relevant properties to verify, as the distribution of possible inputs cannot be formally specified. With Julien Girard-Satabin's PhD thesis, we propose to take advantage of the simulators often used either to train machine learning models or to check them with statistical tests, a growing trend in industry. Our formulation [34] allows us to formally express and verify safety properties on perception units, covering all cases that could ever be generated by the simulator, to the difference of statistical tests which cover only seen examples. Along with this theoretical formulation, we provide a tool to translate deep learning models into standard logical formulae. As a proof of concept, we train a toy example mimicking an autonomous car perceptive unit, and we formally verify that it will never fail to capture the relevant information in the provided inputs.

Experimental validation of Autonomous Vehicle Command Statistical guarantees (e.g., less than 10^{-8} failure per hour of operation) are obtained by empirical tests, involving millions of kilometers of driving in all possible road, weather and traffic conditions as well as intensive simulations, the only way to full control of the driving conditions. The validation process thus involves 3 steps: i) making sure that all parts of the space of possible scenarios are covered by experiments/tests with sufficiently fine grain; ii) identify failures zones in the space of scenarios; iii) fix the controller flaws that resulted in these failures.

TAU is collaborating with Renault on step ii) within Marc Nabhan's CIFRE PhD (defense expected in Sept. 2020). The current target scenario is the insertion of a car on a motorway, the "drosophila" of autonomous car scenarios and the goal is the identification of the conditions of failures of the autonomous car controller. Only simulations are considered here, with one scenario being defined as a parameter setting of the in-house simulator SCANeR. The goal is the detection of as many failures as possible, running as few simulations as possible, and the identification of the borders of the failure zone using an as simple as possible description, thus allowing engineers to understand the reasons for the flaws. A first paper was published [37] proposing several approaches for the identification of failures. On-going work is concerned with a precise yet simple definition of the border of the failure zone.

Reinforcement Learning from Advice In the context of his CIFRE PhD with Thalès, Nizam Makdoud tests (in simulation) physical security systems using reinforcement learning to learn the best sequence of action that will break through the system. This lead him to propose an original approach called *LEarning from Advice* (LEA) that uses knowledge from several policies learned on different tasks. Whereas Learning by imitation uses the actions of the known policy, the proposed method uses the different Q-functions of the known policies. The main advantage of this strategy is its robustness to poor advice, as the policy then reverts to standard DDPG [127]. The results (submitted) demonstrate that LEA is able to learn faster than DDPG if given good-enough policies, and only slightly slower when given lousy advices.

Learning Multi-Criteria Decision Aids (Hierarchical Choquet models) In collaboration with Johanne Cohen (LRI-GALAC) and Christophe Labreuche (Thalès), the representation and data-driven elicitation of hierarchical Choquet models has been tackled. A specific neural architecture, enforcing by design the model constraints (monotonicity, additivity), and supporting the end-to-end training of the Multi-Criteria Decision aid, has been proposed in Roman Bresson's PhD. Under mild assumptions, an identifiability result (existence and unicity of the sought model in the neural space) is obtained. The approach is empirically validated and successfully compared to the state of the art.

7.2. Learning to Learn

*7.2.1. Auto-**

Participants: Guillaume Charpiat, Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhDs: Léonard Blier, Guillaume Doquet, Zhengying Liu, Herilalaina Rakotoarison, Lisheng Sun, Pierre Wolinski
Collaboration: Vincent Renault (SME Artelys); Yann Ollivier (Facebook)

Auto-☆ studies at TAU investigate several research directions.

As mentioned in Section 3.3, a popular approach for algorithm selection is collaborative filtering. In Lisheng Sun's PhD [15], active learning was used on top of the CofiRank algorithm for matrix factorization [164], improving the results and the time to solution of the recommendation algorithm. Furthermore, most real-world domains evolve with time, and an important issue in real-world applications is that of life-long learning, as static models can rapidly become obsolete. Another contribution in Lisheng's PhD is an extension of AutoSklearn that detects concept drifts and corrects the current model accordingly.
An original approach to Auto-☆, explored in Herilalaina Rakotoarison's PhD, extends and adapts Monte-Carlo Tree Search to explore the structured space of pre-processing + learning algorithm configurations, and gradually determine the best pipeline [40]; the resulting MOSAIC algorithm performs on par with AutoSklearn, the winner of Auto-☆ international competitions in the last few years.

Auto-☆ would be much easier if appropriate and affordable meta-features (describing datasets) were available. Taking inspiration from equivariant learning [144] and learning from distributions [130], on-going work aims to learn such meta-features, based on the OpenML archive [161].

A key building block in Auto-☆ lies in the data preparation and specifically variable selection. Guillaume Doquet's PhD [13] addresses the problem of agnostic feature selection, independently of any target variable. The point d'orgue of his work is Agnos [32] (Best Paper Award at ECML 2019), that combines an AutoEncoder with structural regularizations to sidestep the combinatorial optimization problem at the core of feature selection. The extensive experimental validation of AgnoS on the scikit-feature benchmark suite demonstrates its ability compared to the state of the art, both in terms of supervised learning and data compression.
Several works have focused on the adjustment of specific hyper-parameters for neural nets. Pierre Wolinski's PhD (to be defended in January 2020, publication submitted) studies three such hyper-parameters: i) network width (number of neurons in each layer); ii) regularizer importance in the objective function to minimize (factor balancing data term and regularizer); and iii) learning rate. Regarding the network width, it is adjusted during training thanks to a criterion quantifying each neuron's importance, naturally leading to a sparsification effect (as for L1 norm minimization). This study is actually extendable to not only layers' widths but also layers' connectivity (e.g., in modern networks where each layer may be connected to any other layer with 'skip' connections). Regarding the regularizer weight, it is formulated as a probabilistic prior from a Bayesian perspective, which leads to a particular value that the regularizer weight should have in order the network to satisfy some property. Regarding the learning rate, Pierre Wolinski and Leonard Blier [27] proposed to attach fixed learning rates to each neuron (picked randomly) and calibrate this learning rate distribution in such a way that neurons are sequentially active, learning in an optimally agile manner during a first learning phase, and being stable in later phases. This remove the need to tune the learning rate.

A last direction of investigation concerns the design of challenges, that contribute to the collective advance of research in the Auto-☆ direction. The team has been very active in the series of AutoML challenges [154], and steadily contributes to the organization of new challenges (Section 7.6).

7.2.2. Deep Learning: Practical and Theoretical Insights

Participants: Guillaume Charpiat, Marc Schoenauer, Michèle Sebag
PhDs: Léonard Blier, Corentin Tallec
Collaboration: Yann Ollivier (Facebook AI Research, Paris), the Altschuler and Wu lab. (UCSF, USA), Y. Tarabalka (Inria Titane)

Although a comprehensive mathematical theory of deep learning is yet to come, theoretical insights from information theory or from dynamical systems can deliver principled improvements to deep learning and/or explain the empirical successes of some architectures compared to others.

In his PhD [16], Corentin Tallec presents several contributions along these lines:

- In [158], it is shown that the LSTM structure can be understood from axiomatic principles, enforcing the *robustness of the learned model to temporal deformation (warpings) in the data*. The complex LSTM architecture, introduced in the 90's, has become the currently dominant architecture for modeling temporal sequences (such as text) in deep learning. It is shown that this complex LSTM architecture necessarily arises if one wants the model to be able to handle time warpings in the data (such as arbitrary accelerations or decelerations in the signal), and their complex equations can be derived axiomatically.
- In [132] (oral presentation at ICML 2018) the issue of mode dropping in adversarial generative models is tackled using information theory. The adversary (discriminator) task is set to predict the proportion of true and fake images in a set of images, via an information theory criterion, thus working at the level of the *overall distribution* of images. The discriminator is thus mande more able to detect statistical imbalances between the modes created by the generator, thereby reducing the mode dopping phenomenon. The proposed architecture, inspired from equivariant approaches, is provably able to detect all permutation-invariant statistics in a set of images.
- In [159], the problem of recurrent network training is tackled via the theory of dynamical systems by proposing a simple fully online solution avoiding the "time rewind" step, based on real-time, noisy but unbiased approximations of model gradients, which can be implemented easily in a black-box fashion on top of any recurrent model, and which is well-justified mathematically. The price to pay is an increase of variance.
- In [42], we identify sensitivity to time discretization of Deep RL in near continuous-time environments as a critical factor. Empirically, we find that Q-learning-based approaches collapse with small time steps. Formally, we prove that Q-learning does not exist in continuous time. We detail a principled way to build an off-policy RL algorithm that yields similar performances over a wide range of time discretizations, and confirm this robustness empirically.

Several other directions have been investigated: In [41], we introduce a multi-domain adversarial learning algorithm in the semi-supervised setting. We extend the single source H-divergence theory for domain adaptation to the case of multiple domains, and obtain bounds on the average- and worst-domain risk in multi-domain learning. This leads to a new loss to accommodate semi-supervised multi-domain learning and domain adaptation. We obtain state-of-the-art results on two standard image benchmarks, and propose as a new benchmark a novel bioimage dataset, CELL, in the domain of automated microscopy data, where cultured cells are imaged after being exposed to known and unknown chemical perturbations, and in which each dataset displays significant experimental bias.

Another direction regards the topology induced by a trained neural net, and how similar are two samples in the NN perspective. The definition proposed in [29] relies on varying the NN parameters and examining whether the impacts of this variation on both samples are aligned. The mathematical properties of this similarity measure are investigated and the similarity is shown to define a kernel on the input space. This kernel can be used to tractably estimate the sample density, and it leads to new directions for the statistical learning analysis of NN, e.g. in terms of additional loss (requiring that similar examples have a similar latent representation in the above sense) or in terms of resistance to noise. Specifically, a multimodal image registration task is presented where almost perfect accuracy is reached, despite a high label noise (see Section 7.5.2). Such an impressive self-denoising phenomenon can be explained and quantified as a noise averaging effect over the labels of similar examples.

7.2.3. Analyzing and Learning Complex Systems

Participants: Cyril Furtlehner, Aurélien Decelle, François Landes
PhDs: Giancarlo Fissore
Collaboration: Jacopo Rocchi (LPTMS Paris Sud); the Simons team: Rahul Chako (post-doc), Andrea Liu (UPenn), David Reichman (Columbia), Giulio Biroli (ENS), Olivier Dauchot (ESPCI).; Clément Vignax (EPFL); Yufei Han (Symantec).
The information content of a trained restricted Boltzmann machine (RBM) can be analyzed by comparing the singular values/vectors of its weight matrix, referred to as modes, to that of a random RBM (typically following a Marchenko-Pastur distribution) [83]. The analysis of a single learning trajectory is replaced by analyzing the distribution of a well chosen ensemble of models. In G. Fissore's PhD, the learning trajectory of an RBM is shown to start with a linear phase recovering the dominant modes of the data, followed by a non-linear regime where the interaction among the modes is characterized [84]. Although simplifying assumptions are required for a mean-field analysis in closed form of the above distribution, it nevertheless delivers some simple heuristics to speed up the learning convergence and to simplify the models.
This analysis will be extended along two directions: handling missing data [58]; and considering exactly solvable RBM (non-linear RBM for which the contrastive divergence can be computed in closed forms, e.g. using a spherical model) [54]. W.r.t. missing data, state of the art results have been obtained on semi-supervised tasks in the context of Internet-of-Things security, considering a high rate of missing inputs and labels). On the theoretical side exact generic RBM learning trajectories have been characterized, showing intriguing connections based on Bose-Einstein condensation mechanism associated to information storing. Our collaboration with J. Rocchi(LPTMS, Univ. Paris Sud) aims to characterize the landscape of RBMs learned from different initial conditions, and to relate this landscape to the number of parameters (hidden nodes) of the system.

An emerging research topic, concerns the interpretation of deep learning by means of Gaussian processes and associated neural tangent limit kernel in the thermodynamical limit obtained by letting layer's width go to infinity [118]. Various things are planned to be investigated on the basis of this theoretical tool in particular how this translate to RBM or DBM setting, and whether a double dip behaviour is to be expected as well for generative models.

As mentioned earlier, the use of ML to address fundamental physics problems is quickly growing. This leads to some methodological mistakes from newcomers, that have been investigated by Rémi Perrier (2 month internship). One example is the domain of glasses (how the structure of glasses is related to their dynamics), which is one of the major problems in modern theoretical physics. The idea is to let ML models automatically find the hidden structures (features) that control the flowing or non-flowing state of matter, discriminating liquid from solid states. These models can then help identifying "computational order parameters", that would advance the understanding of physical phenomena [19], on the one hand, and support the development of more complex models, on the other hand. More generally, attacking the problem of amorphous condensed matter by novel Graph Neural Networks (GNN) architectures is a very promising lead, regardless of the precise quantity one may want to predict. Currently GNNs are engineered to deal with molecular systems and/or crystals, but not to deal with amorphous matter. This second axis is currently being attacked in collaboration with Clément

Vignac (PhD Student at EPFL), using GNNs. Furthermore, this problem is new to the ML community and it provides an original non-trivial example for engineering, testing and benchmarking explainability protocols.

7.3. Computational Social Sciences

Computational Social Sciences (CSS) is making significant progress in the study of social and economic phenomena thank to the combination of social science theories and new insight from data science. While the simultaneous advent of massive data and unprecedented computational power has opened exciting new avenues, it has also raised new questions and challenges.

Several studies are being conducted in TAU, about labor (labor markets, platform "micro-work", quality of life and economic performance), about nutrition (health, food, and socio-demographic issues), around Cartolabe, a platform for scientific information system and visual querying and around GAMA, a multi-agent based simulation platform.

7.3.1. Labor Studies

Participants: Philippe Caillou, Isabelle Guyon, Michèle Sebag, Paola Tubaro
PhDs: Diviyan Kalainathan, Guillaume Bied, Armand Lacombe
Post-Docs: Saumya Jetley
Engineers: Raphael Jaiswal, Victor Alfonso Naya
Collaboration: Jean-Pierre Nadal (EHESS); Marco Cuturi, Bruno Crépon (ENSAE); Antonio Casilli, Ulrich Laitenberger (Telecom Paris); Odile Chagny (IRES); Alessandro Delfanti (University of Toronto)

A first area of activity of TAU in Computational Social Sciences is the study of labor, from the functioning of the job market, to the rise of new, atypical forms of work in the networked society of internet platforms, and the quality of life at work.
Job markets Two projects deal with the domain of job markets and machine learning. The DATAIA project Vadore, in collaboration with ENSAE and Pôle Emploi, has two goals. First, to improve the recommendation of jobs for applicants (and the recommendation of applicants to job offers). The main originalities in this project are: i) to use both machine learning and optimal transport to improve the recommendation by learning a matching function for past hiring, and then to apply optimal transport-like bias to tackle market congestion (e.g. to avoid assigning many applicants to a same job offer); ii) to use randomized test on micro-markets (AB testing) in collaboration with Pôle Emploi to test the global impact of the algorithms.

The JobAgile project, BPI-PIA contract, coll. EHESS, Dataiku and Qapa, deals with low salary interim job recommendations. A main difference with the Vadore project relies on the high reactivity of the Qapa and Dataiku startups: i) to actually implement AB-testing; ii) to explore related functionalities, typically the recommendation of formations; iii) to propose a visual querying of the job market, using the Cartolabe framework (below).

The platform economy and digital labor

Another topic concerns the digital economy and the transformations of labor that accompany it. One part of the platform economy carries promises of social, not only techno-economic, innovation. If enthusiasms for a new "sharing economy" or "collaborative economy" have progressively faded away, values of decentralization, autonomy, and flatter coordination are still commonly associated to platforms. A conference paper by P. Tubaro studies how events constitute places where actors of the platform economy negotiate values and collectively drive forms of social change [43].

The platform economy and its effects on labor are also linked to the current developments of AI [22]. In collaboration with A.A. Casilli (Telecom ParisTech), P. Tubaro has received funding from the Union Force Ouvrière, from France Stratégie (a Prime Minister's service), and from MSH Paris-Saclay, to map "micro-work" in France (DiPLab project). The term micro-work refers to small, data-related tasks that are performed online against low remunerations, such as tagging objects in images, transcribing bits of text, and recording utterances aloud. Specialized platforms such as Amazon Mechanical Turk, Clickworker and Microworkers recruit online providers to execute these tasks for their clients, mostly for data-intensive production processes. In addition to poor working conditions and low pay, micro-work raises issues in terms of privacy and data protection, insofar as outside providers are entrusted with data that may include personal information [23].

The results of the DiPLab study were published in a report that attracted significant media attention [52], and presented as part of a large event on micro-work at the headquarters of France Stratégie in June 2019. Two articles relying on DiPlab results, one of which was first published as a working paper [62], [61], are now under review.

A joint franco-Canadian grant obtained by P. Tubaro and A. Delfanti (University of Toronto) enabled the creation of an "International Network on Digital Labor", aiming to bring together scholars interested in various forms of digital platform labor. Inauguration of the network involved the organization of two workshops, one in Paris (June 2019) and the other in Toronto (October 2019).

Further research on how digital platforms transform labor practices and affect the very definition of professions is being undertaken by P. Tubaro as part of a two-year (2018-2020) grant from DARES (French Ministry of Labor), in collaboration with O. Chagny of IRES, a union-funded think-tank), and A.A. Casilli (Telecom ParisTech).

A newly-obtained grant from ANR (with A.A. Casilli and U. Laitenberger, Telecom ParisTech) will enable P. Tubaro to further explore the global production networks that link AI developers and producers to data-related work across national boundaries, following outsourcing chains that extend from France to French-speaking African countries, and from Spain and the USA to parts of Latin America. This new project, entitled "The HUman Supply cHain behind smart technologies" (HUSH), will start in January 2020.

7.3.2. Health, food, and socio-demographic relationships

Participants: Philippe Caillou, Michèle Sebag, Paola Tubaro
Post-doc: Ksenia Gasnikova
Collaboration: Louis-Georges Soler, Olivier Allais (INRA)

Another area of activity concerns the relationships between eating practices, socio-demographic features and health.

The Nutriperso project (IRS Univ. Paris-Saclay, coll. INRA, CEA, CNRS, INSERM, Telecom ParisTech and Univ. Paris-Sud) aims to: i) determine the impact of food items on health (e.g., related to T2 diabetes); ii) identify alternative food items, admissible in terms of taste and budget, and better in terms of health; iii) emit personalized food recommendations> One project motivation is the fact that general recommendations (e.g., *Eat 5 fruit and vegetable per day*) are hardly effective on populations at risk. Based on the Kantar database, reporting the food habits of 20,000 households over 20 years, our challenge is to analyze the food purchases at an unprecedented fine-grained scale (at the barcode level), and to investigate the relationship between diets, socio-demographic features, and body mass index (BMI). The challenge also regards the direction of causality; while some diets are strongly correlated to high BMI, the question is to determine whether, e.g., sugar-free sodas are a cause of obesity, or a consequence thereof, or both a cause and a consequence. A main difficulty is the lack of control populations to assess a diet impact. Such a control population could be approximated in the case of the organic diet, showing a statistically significant impact of this diet on the BMI distribution. The question of finding confounders (e.g. based on wealth or education) or "backdoor" variables in under study.

7.3.3. Scientific Information System and Visual Querying

Participants: Philippe Caillou, Michèle Sebag

Engineers: Anne-Catherine Letournel, Jonas Renault
Collaboration: Jean-Daniel Fekete (AVIZ, Inria Saclay)

A third area of activity concerns the 2D visualisation and querying of a corpus of documents. Its initial motivation was related to scientific organisms, institutes or Universities, using their scientific production (set of articles, authors, title, abstract) as corpus. The Cartolabe project started as an Inria ADT (coll. TAO and AVIZ, 2015-2017). It received a grant from CNRS (coll. TAU, AVIZ and HCC-LRI, 2018-2019). Further extensions, as an open-source platform, are under submission at the time of writing.

The originality of the approach is to rely on the content of the documents (as opposed to, e.g. the graph of co-authoring and citations). This specificity allowed to extend Cartolabe to various corpora, such as Wikipedia, Bibliotheque Nationale de France, or the Software Heritage. Cartolabe was also applied in 2019 to the *Grand Debat* dataset: to support the interactive exploration of the 3 million propositions; and to check the consistency of the official results of the *Grand Debat* with the data.

Among its intended functionalities are: the visual assessment of a domain and its structuration (who is expert in a scientific domain, how related are the domains); the coverage of an institute expertise relatively to the general expertise; the evolution of domains along time (identification of rising topics). A round of interviews with beta-user scientists is under way since end 2019. Cartolabe usage raises questions at the crossroad of human-centered computing, data visualization and machine learning: i) how to deal with stressed items (the 2D projection of the item similarities poorly reflects their similarities in the high dimensional document space; ii) how to customize the similarity and exploit the users' feedback about relevant neighborhoods.

7.3.4. *Multi-Agent based simulation framework for social science*

Participants: Philippe Caillou
Collaboration: Patrick Taillandier (INRA), Alexis Drogoul and Nicolas Marilleau (IRD), Arnaud Grignard (MediaLab, MIT), Benoit Gaudou (Université Toulouse 1)

Since 2008, P. Caillou contributes to the development of the GAMA platform, a multi-agent based simulation framework. Its evolution is driven by the research projects using it, which makes it very well suited for social sciences studies and simulations.

The focus of the development team in 2019 was on the stability of the platform and on the documentation to provide a stable and well documented framework to the users.

7.4. Energy Management

7.4.1. *Power Grids Daily Management*

Participants: Isabelle Guyon, Marc Schoenauer
PhDs: Benjamin Donnot, Balthazar Donon
Collaboration: Antoine Marot, Patrick Panciatici (RTE), Olivier Teytaud (Facebook)

Benjamin Donnot's CIFRE PhD with RTE [12] dealt with Power Grid safety: The goal is to assess in real time the so-called "(n-1) safety" (see Section 4.2) of possible recovery actions modifying the topology of the grid after some problem occurred somewhere on the grid. However, the HADES simulator, that allows to compute the power flows in the whole network, is far too slow to simulate in real time all n-1 possible failures of a tentative topology. A simplified simulator is also available, but its accuracy is too poor to give good results. Deep surrogate models can be trained off-line for a given topology, based on the results of the slow simulator, with high-enough accuracy, but training as many models as possible failures (i.e., n-1) obviously doesn't scale up: the topology of the grid must be an input of the learned model, allowing to instantly compute the power flows at least for grid configurations close to the usual running state of the grid. A standard approach is the one-hot encoding of the topology, where n additional boolean inputs are added to the neural network, encoding the presence or absence of each line. Nevertheless, this approach poorly generalize to topologies outside the distribution of the ones used for training.

An original "guided dropout" approach was first proposed [87], in which the topology directly acts on the connections of the deep network: a missing line suppresses some connections. Whereas the standard dropout method disconnect random connections for every batch, in order to improve the generalization capacity of the network, the "guided dropout" method removes some connections based on the actual topology of the network. This approach is experimentally validated against the one-hot encoding on small subsets of the French grid (up to 308 lines). Interestingly, and rather surprisingly, even though only examples with a single disconnected line are used in the training set, the learned model is able of some additive generalization, and predictions are also accurate enough in the case 2 lines are disconnected. The guided dropout approach was later robustified [86] by learning to rapidly rank higher order contingencies including all pairs of disconnected lines, in order to prioritize the cases where the slow simulator is run: Another neural network is trained to rank all (n-1) and (n-2) contingencies in decreasing order of presumed severity.

The guided dropout approach has been further extended and generalized with the LEAP (Latent Encoding of Atypical Perturbation) architecture [30], [17], by crossing-out connections between the encoder and the decoder parts of the ResNet architecture. LEAP then performs transfer learning over spaces of distributions of topology perturbations, allowing to better handle more complex actions on the tolology, going beyond (n-1) and (n-2) perturbations by also including node-split, a current action in the real world. The LEAP approach was theoretically studied in the case of additive perturbations, and experimentally validated on an actual sub-grid of the French grid with 46 consumption nodes, 122 production nodes, 387lines and 192 substations.

LEAP is also the firs part of Balthazar Donon's on-going PhD, that currently develops using a completely different approach to approximate the power flows on a grid, i.e. that of Graph Neural Networks (GNNs). From a Power Grid perspective, GNNs can be viewed as including the topology in the very structure of the neural network, and learning some generic transfer function amongst nodes that will perform well on any topology. First results [31] use a loss based on a large dataset of actual power flows computed using the slow HADES simulator. The results indeed generalize to very different topologies than the ones used for training, in particular very different sizes of power grids. On-going work [56] removes the need to run HADES thanks to a loss that directly aims to minimize Kirshoff's law on all lines.

7.4.2. Local Grids Optimization, and the Modeling of Worst-case Scenarios

Participants: Isabelle Guyon, Marc Schoenauer, Michèle Sebag
PhDs: Victor Berger, Herilalaina Rakotoarison; **Post-doc**: Berna Batu
Collaboration: Vincent Renaut (Artelys)

One of the goals of the ADEME Next project, in collaboration with SME Artelys (see also Section 4.2), is the sizing and capacity design of regional power grids. Though smaller than the national grid, regional and urban grids nevertheless raise scaling issues, in particular because many more fine-grained information must be taken into account for their design and predictive growth.

Regarding the design of such grids, and provided accurate predictions of consumption are available (see below), off-the-shelf graph optimization algorithms can be used. Berna Batu is gathering different approaches. Herilalaina Rakotoarison's PhD tackles the automatic tuning of their parameters (see Section 7.2.1); while the Mosaic algorithm is validated on standard AutoML benchmarks [40], its application to Artelys' home optimizer at large Knitro is on-going, and compared to the state-of-the-art in parameter tuning (confidential deliverable).

In order to get accurate consumption predictions, V. Berger's PhD tackles the identification of the peak of energy consumption, defined as the level of consumption that is reached during at least a given duration with a given probability, depending on consumers (profiles and contracts) and weather conditions. The peak identification problem is currently tackled using Monte-Carlo simulations based on consumer profile- and weather-dependent individual models, at a high computational cost. The challenge is to exploit individual models to train a generative model, aimed to sampling the collective consumption distribution in the quantiles with highest peak consumption. The concept of *Compositional Variational Auto-Encoder* was proposed: it is amenable to multi-ensemblist operations (addition or subtraction of elements in the composition), enabled by the invariance and generality of the whole framework w.r.t. respectively, the order and number of the elements. It has been first tested on synthetic problems [26].

7.5. Data-driven Numerical Modelling

7.5.1. High Energy Physics

Participants: Cécile Germain, Isabelle Guyon
PhD: Victor Estrade, Adrian Pol
Collaboration: D. Rousseau (LAL), M. Pierini (CERN)

The role and limits of simulation in discovery is the subject of V. Estrade's PhD, specifically uncertainty quantification and calibration, that is how to handle the systematic errors, arising from the differences ("known unknowns") between simulation and reality, coming from uncertainty in the so-called nuisance parameters. In the specific context of HEP analysis, where relatively numerous labelled data are available, the problem is at the crosspoint of domain adaptation and representation learning. We have investigated how to directly enforce the invariance w.r.t. the nuisance in the sought embedding through the learning criterion (tangent back-propagation) or an adversarial approach (pivotal representation). The results [93] contrast the superior performance of incorporating a priori knowledge on a well separated classes problem (MNIST data) with a real case setting in HEP, in relation with the Higgs Boson Machine Learning challenge [68] and the TRrackML challenge [46]. More indirect approaches based on either incorporating variance reduction for the parameter of interest or constraining the representation in a variational auto-encoder farmework are currently considered.

Anomaly detection (AD) is the subject of A. Pol's PhD. Reliable data quality monitoring is a key asset in delivering collision data suitable for physics analysis in any modern large-scale high energy physics experiment. [21] focuses on supervised and semi-supervised methods addressing the identification of anomalies in the data collected by the CMS muon detectors. The combination of DNN classifiers capable of detecting the known anomalous behaviors, and convolutional autoencoders addressing unforeseen failure modes has shown unprecedented efficiency. The result has been included in the production suite of the CMS experiment at CERN. Recent work has focused on improving AD for the trigger system, which is the first stage of event selection process in most experiments at the LHC at CERN. The hierarchical structure of the trigger process called for exploiting the advances in modeling complex structured representations that perform probabilistic inference effectively, and specifically variational autoencoders. Previous works argued that training VAE models only with inliers is insufficient and the framework should be significantly modified in order to discriminate the anomalous instances. In this work, we exploit the deep conditional variational autoencoder (CVAE) and we define an original loss function together with a metric that targets hierarchically structured data AD [39], [64]. This results in an effective, yet easily trainable and maintainable model.

The highly visible TrackML challenge [46] is described in section 7.6.

7.5.2. Remote Sensing Imagery

Participants: Guillaume Charpiat
Collaboration: Yuliya Tarabalka, Armand Zampieri, Nicolas Girard, Pierre Alliez (Titane team, Inria Sophia-Antipolis)

The analysis of satellite or aerial images has been a long-time ongoing topic of research, but the remote sensing community moved only very recently to a principled vision of the tasks in a machine learning perspective, with sufficiently large benchmarks for validation. The main topics are the segmentation of (possibly multispectral) remote sensing images into objects of interests, such as buildings, roads, forests, etc., and the detection of changes between two images of the same place taken at different moments. The main differences with classical computer vision is that images are large (covering whole countries, typically cut into 5000×5000 pixels tiles), containing many small, potentially similar objects (and not one big object per image), that every pixel needs to be annotated (w.r.t. assigning a single label to a full image), and that the ground truth is often not reliable (spatially mis-registered, missing new constructions).

These last years, deep learning techniques took over classical approaches in most labs, adapting neural network architectures to the specifics of the tasks. This is due notably to the creation of several large scale benchmarks (including one by us [133] and, soon after, larger ones by GAFAM).

This year, we continued the work started in [167] about the registration of remote sensing images (RGB pictures) with cadastral maps (made of polygons indicating buildings and roads). We extended it in [33] to the case of real datasets, i.e. to noisy data. Indeed, in remote sensing, datasets are often large but of poor ground truth annotation quality. It turns out that, when training on datasets with noisy labels, one can still obtain accuracy scores far better than the noise variance in the training set, due to averaging effects over the labels of similar examples. To properly explain this, a theoretical study was conducted (cf. Section 7.2.2). Given any already trained neural network and its noisy training set, without knowing the real ground truth, we were then able to quantify this noise averaging effect [29].

We also tackled the problem of pansharpening, i.e. the one of producing a high-resolution color image, given a low-resolution color image and a high-resolution greyscale one [35], with deep convolutional neural networks as well.

7.5.3. *Space Weather Forecasting*

Participants: Cyril Furtlehner, Michèle Sebag
PhD: Mandar Chandorkar
Collaboration: Enrico Camporeale (CWI)

Space Weather is broadly defined as the study of the relationships between the variable conditions on the Sun and the space environment surrounding Earth. Aside from its scientific interest from the point of view of fundamental space physics phenomena, Space Weather plays an increasingly important role on our technology-dependent society. In particular, it focuses on events that can affect the performance and reliability of space-borne and ground-based technological systems, such as satellite and electric networks that can be damaged by an enhanced flux of energetic particles interacting with electronic circuits.[0]

Since 2016, in the context of the Inria-CWI partnership, a collaboration between TAU and the Multiscale Dynamics Group of CWI aims to long-term Space Weather forecasting. The goal is to take advantage of the data produced everyday by satellites surveying the sun and the magnetosphere, and more particularly to relate solar images and the quantities (e.g., electron flux, proton flux, solar wind speed) measured on the L1 libration point between the Earth and the Sun (about 1,500,000 km and 1 hour time forward of Earth). A challenge is to formulate such goals in terms of supervised learning problem, while the "labels" associated to solar images are recorded at L1 (thus with a varying and unknown time lag). In essence, while typical ML models aim to answer the question *What*, our goal here is to answer both questions *What* and *When*. This project has been articulated around Mandar Chandorkar's Phd thesis[11] which has been defended this year in Eindhoven. One of the main result that has been obtained concerns the prediction of solar wind impacting earth magnetosphere from solar images. In this context we encountered an interesting sub-problem related to the non deterministic travel time of a solar eruption to earth's magnetosphere. We have formalized it as the joint regression task of predicting the magnitude of signals as well as the time delay with respect to their driving phenomena. We have provided in[28] an approach to this problem combining deep learning and an original Bayesian forward attention mechanism. A theoretical analysis based on linear stability has been proposed to put this algorithm on firm ground. From the practical point of view, encouraging tests have been performed both on synthetic data and real data with results slightly better than those present in the specialized literature on a small dataset. Various extension of the method, of the experimental tests and of the theoretical analysis are planned.

7.5.4. *Genomic Data and Population Genetics*

Participants: Guillaume Charpiat, Flora Jay, Aurélien Decelle, Cyril Furtlehner
PhD: Théophile Sanchez – **PostDoc**: Jean Cury

[0] After a recent survey conducted by the insurance company Lloyd's, an extreme Space Weather event could produce up to $2.6 trillion in financial damage.

Collaboration: Bioinfo Team (LRI), Estonian Biocentre (Institute of Genomics, Tartu, Estonia), Pasteur Institute (Paris), TIMC-IMAG (Grenoble)

Thanks to the constant improvement of DNA sequencing technology, large quantities of genetic data should greatly enhance our knowledge about evolution and in particular the past history of a population. This history can be reconstructed over the past thousands of years, by inference from present-day individuals: by comparing their DNA, identifying shared genetic mutations or motifs, their frequency, and their correlations at different genomic scales. Still, the best way to extract information from large genomic data remains an open problem; currently, it mostly relies on drastic dimensionality reduction, considering a few well-studied population genetics features.

We developed an approach that extracts features from genomic data using deep neural networks and combines them with a Bayesian framework to approximate the posterior distribution of demographic parameters. The key difficulty is to build flexible problem-dependent architectures, supporting transfer learning and in particular handling data with variable size. We designed new generic architectures, that take into account DNA specificities for the joint analysis of a group of individuals, including its variable data size aspects and compared their performances to state-of-the-art approaches [148]. In the short-term these architectures can be used for demographic inference or selection inference in bacterial populations (ongoing work with a postdoctoral researcher, J Cury, and the Pasteur Institute); the longer-term goal is to integrate them in various systems handling genetic data or other biological sequence data.

In collaboration with the Institute of Genomics of Tartu (Estonia; B Yelmen, 3-month visitor at LRI), we leveraged two types of generative neural networks (Generative Adversarial Networks and Restricted Boltzmann Machines) to learn the high dimensional distributions of real genomic datasets and create artificial genomes [66]. These artificial genomes retain important characteristics of the real genomes (genetic allele frequencies and linkage, hidden population structure, ...) without copying them and have the potential to be valuable assets in future genetic studies by providing anonymous substitutes for private databases (such as the ones hold by companies or public institutes like the Institute of Genomics of Tartu). Yet, ensuring anonymity is a challenging point and we measured the privacy loss by using and extending the Adversarial Accuracy score developed by the team for synthetic medical data [44].

In collaboration with TIMC-IMAG, we proposed a new factor analysis approach that process genetic data of multiple individuals from present-day and ancient populations to visualize population structure and estimate admixture coefficients (that is, the probability that an individual belongs to different groups given the genetic data). This method corrects the traditionally-used PCA by accounting for time heterogeneity and enables a more accurate dimension reduction of paleogenomic data [59].

7.5.5. *Sampling molecular conformations*

Participants: Guillaume Charpiat
PhD: Loris Felardos
Collaboration: Jérôme Hénin (IBPC), Bruno Raffin (InriAlpes)

Numerical simulations on massively parallel architectures, routinely used to study the dynamics of biomolecules at the atomic scale, produce large amounts of data representing the time trajectories of molecular configurations, with the goal of exploring and sampling all possible configuration basins of given molecules. The configuration space is high-dimensional (10,000+), hindering the use of standard data analytics approaches. The use of advanced data analytics to identify intrinsic configuration patterns could be transformative for the field.

The high-dimensional data produced by molecular simulations live on low-dimensional manifolds; the extraction of these manifolds will enable to drive detailed large-scale simulations further in the configuration space. This year, we studied how to bypass simulations by directly predicting, given a molecule formula, its possible configurations. This is done using Graph Neural Networks [89] in a generative way, producing 3D configurations. The goal is to sample all possible configurations, and with the right probability.

7.5.6. Storm trajectory prediction

Participants: Guillaume Charpiat
Collaboration: Sophie Giffard-Roisin (IRD), Claire Monteleoni (Boulder University), Balazs Kegl (LAL)

Cyclones, hurricanes or typhoons all designate a rare and complex event characterized by strong winds surrounding a low pressure area. Their trajectory and intensity forecast, crucial for the protection of persons and goods, depends on many factors at different scales and altitudes. Additionally storms have been more numerous since the 1990s, leading to both more representative and more consistent error statistics.

Currently, track and intensity forecasts are provided by numerous guidance models. Dynamical models solve the physical equations governing motions in the atmosphere. While they can provide precise results, they are computationally demanding. Statistical models are based on historical relationships between storm behavior and other parameters [82]. Current national forecasts are typically driven by consensus methods able to combine different dynamical models.

Statistical models perform poorly compared to dynamical models, although they rely on steadily increasing data resources. ML methods have scarcely been considered, despite their successes in related forecasting problems [169]. A main difficulty is to exploit spatio-temporal patterns. Another difficulty is to select and merge data coming from heterogeneous sensors. For instance, temperature and pressure are real values on a 3D spatial grid, while sea surface temperature or land indication rely on a 2D grid, wind is a 2D vector field, while many indicators such as geographical location (ocean, hemisphere...) are just real values (not fields), and displacement history is a 1D vector (time). An underlying question regards the *innate* vs *acquired* issue, and how to best combine physical models with trained models. The continuation of the work started last year [101] shows that with deep learning one can outperform the state-of-the-art in many cases [18].

7.6. Challenges

Participants: Cécile Germain, Isabelle Guyon, Adrien Pavao, Anne-Catherine Letournel, Michèle Sebag
PhD: Zhengying Liu, Lisheng Sun, Balthazar Donon
Collaborations: D. Rousseau (LAL), André Elisseeff (Google Zurich), Jean-Roch Vilmant (CERN), Antoine Marot and Benjamin Donnot (RTE), Kristin Bennett (RPI), Magali Richard (Université de Grenoble).

The TAUgroup uses challenges (scientific competitions) as a means of stimulating research in machine learning and engage a diverse community of engineers, researchers, and students to learn and contribute advancing the state-of-the-art. The TAUgroup is community lead of the open-source **Codalab** platform, hosted by Université Paris-Saclay. The project had grown in 2019 and includes now an engineer dedicated full time to administering the platform and developing challenges (Adrien Pavao), financed by a new project just starting with the Région Ile-de-France. This project will also receive the support of the Chaire Nationale d'Intelligence Artificielle of Isabelle Guyon for the next four years.

Following the highly successful ChaLearn **AutoML** Challenges (NIPS 2015 – ICML 2016 [111] – PKDD 2018 [113]), a series of challenges on the theme of **AutoDL** [129] was run in 2019 (see http://autodl.chalearn.org, addressing the problem of tuning the hyperparameters of Deep Neural Networks, including the topology of the network itself. Co-sponsored by Google Zurich, it required participants to upload their code on the Codalab platform. The series included two challenges in computer vision called **AutoCV** and **AutoCV2**, to promote automatic machine learning for image and video processing, in collaboration with University of Barcelona [45]. It also included challenges in speech processing (**AutoSpeech**), text processing (**AutoNLP**), weakly supervised learning (**AutoWeakly**) and times series (**AutoSeries**), co-organized with 4Paradigm. It culminated with launching the **AutoDL** challenge combining multiple modalities (presently on-going). The winners of each challenge open-sourced their code. GPU cloud resources were donated by Google. AutoDL was an official NeurIPS 2020 competition.

Part of the High Energy Physics activities of the team, **TrackML** [79], [80] first phase was run and co-sponsored by Kaggle, until September 2018. The second phase has been run on Codalab until March 2019, requiring code submission; algorithms were then ranked by combining accuracy and speed. The best submissions largely outperform the existing solutions. The challenge has been presented at NeurIPS [46], and at a CERN workshop[0]. I. Guyon and C. Germain are in the organizing committee, and M. Schoenauer is member of the Advisory Committee. The TAU team, in collaboration with CERN, has taken a leading role in stimulating both the ML and HEP communities to address the combinatorial complexity explosion created by the next generation of particle detectors.

A new challenge series in Reinforcement Learning was started with the company RTE France, one the theme "Learning to run a power network" [134] (**L2RPN**, http://l2rpn.chalearn.org). The goal is to test the potential of Reinforcement Learning to solve a real world problem of great practical importance: controlling electricity transportation in smart grids while keeping people and equipment safe. The first edition was run in Spring 2019 and was part of the official selection of the IJCNN 2019 conference. It ran on the Codalab platform coupled with the open source PyPower simulator of power grids interfaced with the Opengym RL framework, developed by OpenAI. In this gamified environment, the participants had to create a proper controller of a small grid of 14 nodes. Not all of them used RL, but some combinations of RL and human expertise proved to be competitive. In 2020, we will launch a new edition of the challenge with a more powerful simulator rendering the grid more realistic and capable of simulating a 118-node grid within our computational constraints. This competition was already accepted as part of the official program of IJCNN 2020.

The **HADACA** project (EIT Health) aims to run a series of challenges to promote and encourage innovations in data analysis and personalized medicine. Université de Grenoble organized a challenge on matrix factorization (https://www.medinfo-lyon.org/en/matrixen) using Codalab. The challenge gathered transdisciplinary instructors (researchers and professors), students, and health professionals (clinicians). The HADACA project contributed to create a large dataset to assess tumor heterogeneity in cancer research as well as developing innovative pedagogical methods to sensitize students to big data analysis in health. One of the products of HADACA is the ChaGrade platform (https://chagrade.lri.fr/), a tool allowing instructors to easily use challenges in the classroom, grading them as homework, and monitoring submissions and progress. HADACA will be pursued in 2020 by a sequel project also funded by EIT Health, called COMETH. The objective of COMETH will be to create an environment to conduct systematic benchmarks, based on Codalab. As a synergistic activity, TAU is also engaged in a collaboration with the Rensselaer Polytechnic Institute (RPI, New-York, USA) to use challenges in the classroom, as part of their health-informatics curriculum.

It is important to introduce **challenges in ML teaching**. This has been done (and is on-going) in I. Guyon's Licence and Master courses [38] : some assignments to Master students are to design small challenges, which are then given to Licence students in labs, and both types of students seem to love it. Codalab has also been used to implement reinforcement learning homework in the form of challenges by Victor Berger and Heri Rakotoarison for the class of Michèle Sebag. Along similar line, F. Landes proposed a challenge in the context of S. Mallat's course, at Collège de France. Finally, in collaboration with aiforgood.org, and Heri Rakotoarison has put in place a hackathon for the conference Data Science Africa (https://codalab.lri.fr/competitions/522)

In terms of dissemination, four books were published in 2019 in the Springer series on challenges in machine learning, see http://www.chalearn.org/books.html.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

TAU will continue TAO policy about technology transfer, accepting any informal meeting following industrial requests for discussion (and we are happy to be too much solicited), and deciding about the follow-up based upon the originality, feasibility and possible impacts of the foreseen research directions, provided they fit our

[0] https://indico.cern.ch/event/813759/

general canvas. This lead to the following 5 on-going CIFRE PhDs, with the corresponding side-contracts with the industrial supervisor, plus 3 other bilateral contracts. In particular, we now have a first “Affiliate” partner, the SME DMH, and hope to further develop in the future this form of transfer. Note that it can also sometimes lead to collaborative projects, as listed in the following sections.

- **DMH** 2019 (1 an, 45kEuros) related to consulting activities with DMH (Digital for Mental Health)[0].
  Coordinator: Aurélien Decelle and Simon Moulieras (DMH)
  Participants: Michèle Sebag
- **CIFRE Renault** 2017-2020 (45 kEuros), related to Marc Nabhan’s CIFRE PhD *Sûreté de fonctionnement d’un véhicule autonome - évaluation des fausses détections au travers d’un profil de mission réduit*
  Coordinator: Marc Schoenauer and Hiba Hage (Renault)
  Participants: Marc Nabhan (PhD), Yves Tourbier (Renault)
- **BOBCAT** The new B-tO-B work intermediaries: comparing business models in the "CollaborATive" digital economy, 2018-2020 (100k euros), funded by DARES (French Ministry of Labor).
  Coordinator : Odile Chagny (IRES)
  Participants: Paola Tubaro and Antonio A. Casilli (Telecom Paris)
- **INDL-KW** International Network on Digital Labor - Kickoff Workshops, 2019 (10k euros), funded by CNRS and the University of Toronto.
  Coordinator: Paola Tubaro and Alessandro Delfanti (UoT)
  Participants: Antonio A. Casilli (Telecom Paris)
- **CIFRE Thalès** 2018-2021 (45 kEuros), with Thales Teresis, related to Nizam Makdoud’s CIFRE PhD
  Coordinator: Marc Schoenauer and Jérôme Kodjabatchian
  Participants: Nizam Makdoud
- **CIFRE RTE** 2018-2021 (72 kEuros), with Réseau Transport d’Electricité, related to Balthazar Donon’s CIFRE PhD
  Coordinator: Isabelle Guyon and Antoine Marot (RTE)
  Participants: Balthazar Donon, Marc Schoenauer
- **CIFRE FAIR** 2018-2021 (45 kEuros), with Facebook AI Research, related to Leonard Blier’s CIFRE PhD
  Coordinator: Marc Schoenauer and Yann Olliver (Facebook)
  Participants: Guillaume Charpiat, Michèle Sebag, Léonard Blier
- **IFPEN** (Institut Français du Pétrole Energies Nouvelles) 2019-2023 (300 kEuros), to hire an Inria Starting Research Position (PhD + 4-6 years) to work in all topics mentioned in Section 3.2 relevant to IFPEN activity (see also Section 4.2). Started October 2019.
  Coordinator: Marc Schoenauer
  Participants: Alessandro Bucci, Guillaume Charpiat

[0]This “Affiliate” contract has been inspired by the affiliate program of Technion

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

- **EPITOME** 2017-2020 (225kEuros), *Efficient rePresentatIon TO structure large-scale satellite iMagEs* (Section 7.5.2).
  Coordinator: Yuliya Tarabalka (Titane team, Inria Sophia-Antipolis)
  Participant: Guillaume Charpiat
- **HUSH** 2020-2023 (348kEuros), *The HUman Supply cHain behind smart technologies*.
  Coordinator: Antonio A. Casilli (Telecom Paris)
  Participant: Paola Tubaro

9.1.2. Others

- **Nutriperso** 2017-2020, 122 kEuros. Personalized recommendations toward healthier eating practices (Section 7.3.2).
  U. Paris-Saclay IRS (*Initiative de Recherche Stratégique*)
  Partners: INRA (coordinator), INSERM, Agro Paristech, Mines Telecom
  Participants: Philippe Caillou, Flora Jay, Michèle Sebag, Paola Tubaro
- **IRS CDS** 2017-2020, 75 kEuros. Personalized recommendations toward healthier eating practices
  U. Paris-Saclay IRS (*Initiative de Recherche Stratégique*)
  Partners: INRA (coordinator), INSERM, Agro Paristech, Mines Telecom
  Participants: Philippe Caillou, Flora Jay, Michèle Sebag, Paola Tubaro
- **PIA Adamme** 2015-2019 (258 kEuros) Machine Learning on a mass-memory architecture.
  Coordinator: Bruno Farcy (Bull SAS)
  Participants: Marc Schoenauer, Guillaume Charpiat, Cécile Germain-Renaud
- **NEXT** 2017-2021 (675 kEuros). Simulation, calibration, and optimization of regional or urban power grids (Section 4.2).
  ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie)
  Coordinator: ARTELYS
  Participants Isabelle Guyon, Marc Schoenauer, Michèle Sebag, Victor Berger (PhD), Herilalaina Rakotoarison (PhD), Berna Bakir Batu (Post-doc)
- **DATAIA Vadore** 2018-2020 (105 kEuros) VAlorizations of Data to imprOve matching in the laboR markEt, with CREST (ENSAE) and Pôle Emploi (Section 7.3.1).
  Coordinator: Michèle Sebag
  Participants: Philippe Caillou, Isabelle Guyon
- **PIA JobAgile** 2018-2021 (379 kEuros) *Evidence-based Recommandation pour l'Emploi et la Formation* (Section 7.3.1).
  Coordinator: Michèle Sebag and Stéphanie Delestre (Qapa)
  Participants: Philippe Caillou, Isabelle Guyon
- **HADACA** 2018-2019 (50 kEuros), within EIT Health, for the organization of challenges toward personalized medicine (Section 7.6).
  Coordinator: Magali Richard (Inria Grenoble)
  Participants: Isabelle Guyon
- **IPL HPC-BigData** 2018-2022 (100 kEuros) High Performance Computing and Big Data (Section 7.5.5)
  Coordinator: Bruno Raffin (Inria Grenoble)
  Participants: Guillaume Charpiat, Loris Felardos (PhD)

- **ScGlass** 2016-2020 (10 M$), “Cracking the Glass problem” international collaboration on cracking the glass problem, funded by the Simons Fundation (NY, NYC, USA).
  Coordinator: 13 PIs around the world (see https://scglass.uchicago.edu/)
  Participants: (alumni, actively collaborating with members) François Landes

9.2. European Initiatives

9.2.1. Collaborations with Major European Organizations

CERN: collaboration with two major CERN experiments (ATLAS and CMS) on the role of machine learning at all stages of the scientific discovery process. C. Germain supervises a CERN-funded PhD.

9.3. International Initiatives

9.3.1. Inria International Labs

IIL CWI-Inria

Associate Team involved in the International Lab:

9.3.1.1. MDG-TAO

Title: Data-driven simulations for Space Weather predictions

International Partner (Institution - Laboratory - Researcher):

CWI (Netherlands) - Multiscale Dynamics Group - Enrico Camporeale

Start year: 2017

See also: http://pages.saclay.inria.fr/cyril.furtlehner/html/mdg-tao.html

We propose an innovative approach to Space Weather modeling: the synergetic use of state-of-the-art simulations with Machine Learning and Data Assimilation techniques, in order to adjust for errors due to non-modeled physical processes, and parameter uncertainties. We envision a truly multidisciplinary collaboration between experts in Computational Science and Data assimilation techniques on one side (CWI), and experts in Machine Learning and Data Mining on the other (Inria). Our research objective is to realistically tackle long-term Space Weather forecasting, which would represent a giant leap in the field. This proposal is extremely timely, since the huge amount of (freely available) space missions data has not yet been systematically exploited in the current computational methods for Space Weather. Thus, we believe that this work will result in cutting-edge results and will open further research topics in space Weather and Computational Plasma Physics.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

Isabelle Guyon - Competition co-chair, ECMLPKDD 2019

Michele Sebag - Area Chair NIPS 2017-2019, ICLR 2020

Flora Jay - PASADENA workshop co-chair, Paris, 2019

10.1.1.2. Member of Organizing Committees

Guillaume Charpiat - Organizing & scientific committee of the ForMaL summer school, at ENS Cachan, June 2019, and of WAISE (Second International Workshop on Artificial Intelligence Safety Engineering) held at SafeComp in September 2019.

Flora Jay - Organizing & scientific committee of Research Program "Ecosystem dynamics : stakes, data and models" at Institut Pascal, Paris-Saclay, 2019.

Isabelle Guyon - Advisory committee BayLearn 2019; Co-organizer NeurIPS 2019 workshop on Challenges in Machine Learning; Co-organizer NeurIPS 2019 NewInML workshop.

Marc Schoenauer - Steering Committee, Parallel Problem Solving from Nature (PPSN); Steering Committee, Learning and Intelligent OptimizatioN (LION).

Michele Sebag - President of Steering Committee, Eur. Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD).

10.1.2. Scientific Events: Selection

10.1.2.1. Reviewer

All TAU members are reviewers of the main conferences in their respective fields of expertise.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Isabelle Guyon - Action editor, *Journal of Machine Learning Research* (JMLR); series editor, *Springer series Challenges in Machine Learning* (CiML).

Marc Schoenauer - Advisory Board, *Evolutionary Computation Journal*, MIT Press, and *Genetic Programming and Evolutionary Machines*, Springer Verlag; Action editor, *Journal of Machine Learning Research* (JMLR).

Michèle Sebag - Editorial Board, Machine Learning, Springer Verlag.

Paola Tubaro - Associate Editorial Board *Sociology*, Sage; Co-editor, *Revue Française de Sociologie*, Presses de Sciences Po.

10.1.3.2. Reviewer - Reviewing Activities

All members of the team reviewed numerous articles for the most prestigious journals in their respective fields of expertise.

10.1.4. Invited Talks

Guillaume Charpiat - *Deep Learning for Satellite Imagery* in the Imagine team (ENPC), at Champs-sur-Marne, March 2019; and at LRDE lab (EPITA), at Kremlin-Bicêtre, April 2019; *Deep Learning for Storm Trajectory Prediction and Remote Sensing* at the seminar AI for Climate, at Jussieu (Paris), December 2019.

Cyril Furtlehner - *A machine learning approach to solar wind speed forecasting from solar images*, Machine Learning in Heliophysics conference, Amsterdam September 2019.

Flora Jay - *Machine Learning and Deep Learning for Population Genetics*, Statistics/Machine Learning at Paris-Saclay, Bures-sur-Yvette, Jan 2019; *Inferring Demography? A Deep Learning Approach for Population Genetic Data*, ALPHY 7-8 Feb. 2019; *Génomes présents, histoires d'antan: Apprentissage profond pour la génétique des populations*, Inria Seminar Unithé ou Café with G Charpiat, Feb 2019; *When AI & Big Data meet life sciences - Advances in research and ethical questions*, Round Table at YLRS2019, Paris, Jun 2019; *Inferring past history from genetic data using ABC and Deep Learning approaches*, Seminar at Lille University, Jan 2019; *Creating Artificial Human Genomes Using Generative Models*, Seminars at LCQB and Evolmol IBENS, Paris, Nov and Dec 2019.

Julien Girard - *Formal validation for machine learning*, DataIA day on Safety & AI, Palaiseau, Sept. 11th.

Isabelle Guyon - Invited talk at 2019 Sackler Coloquium of the Science of Deep Learning, and Keynote IJCNN 2019: "Neural network solvers for power transmission problems" (https://youtu.be/YBcRzIAFDYU).

Michèle Sebag - *Meta-Learning*, Kickoff of the Kompetenz Center in Machine Learning, Rhine Ruhr Region, Jan. 2019; *Structural Agnostic Models*, Oberwolfach Symposium on Causality, May 2019; *Artificial Intelligence & Causal Modeling*, CREST Big Data Applications Symposium, Tokyo, Sept. 24, 2019; *Some news and questions about AI and Machine Learning*, Arensberg Symposium, Leuven, Nov. 2019.

Marc Schoenauer - *Intelligence Artificielle Mythes et réalités*, audition par la Fédération Française du Bâtiment, Feb. 13, 2019; *Intelligence Artificielle : du congrès de Dartmouth au rapport Villani*, Séminaire IFPEN, Apr. 4. 2019; *Intelligence Artificielle : de Dartmouth à l'apprentissage profond et au rapport Villani*, Rencontres Franciliennes de Mécanique, May 28, 2019; *A gentle introduction to AI and (Deep) Learning*, Mexican-French Workshop, Aug. 27. 2019, Mexico; *Intelligence artificielle : certification, transparence, et impact sur la société*, Data Science Day, Mines Paris Tech, Sept. 18, 2019; *When Big Data and Machine Learning meet Partial Differential Equations*, CREST Big Data Applications Symposium, Tokyo, Sept. 25, 2019;

Paola Tubaro - *Sélectionné.e par une IA ? Algorithmes, inégalités, et les "humains dans la boucle"*, Centre D'Alembert, Orsay, 18/04/2019; *Dans la fabrique des algorithmes : plateformes, micro-travail et dynamiques d'externalisation*, INSEE, 14/05/2019; *The human labour that makes AI possible: An empirical study of micro-work in France*, Alan Turing Institute, London, 22/01/2019; *Que font les big data aux sciences sociales ? Retour sur une 'crise' annoncée*, EHESS, 21/02/2019.

10.1.5. Leadership within the Scientific Community

Isabelle Guyon - President and co-founder of ChaLearn, a non-for-profit organization dedicated to the organization of challenge.

Marc Schoenauer - Chair of ACM-SIGEVO (Special Interest Group on Evolutionary Computation), 2015-2019, now in Advisory Board; Founding President (since 2015) of SPECIES (Society for the Promotion of Evolutionary Computation In Europe and Surroundings), that organizes the yearly series of conferences *EvoStar*.

Michèle Sebag - Elected Chair of Steering Committee, ECML-PKDD; Board member, Institut de Convergence DataIA.

Paola Tubaro - Convenor of the Social Network Analysis Group of British Sociological Association; co-founder of European Network on Digital Labor.

10.1.6. Scientific Expertise

Guillaume Charpiat - Member of the Inria Saclay Commission Scientifique, and as such, jury committee for grants for PhD theses / postdocs / professor delegations; expertise for DigiCosme grants, ANRT CIFRE PhD grants; for GPU platforms: Jean Zay (GENCI) and Lab-IA (Saclay plateau). Discussion panel of the workshop day IA & Océan-Atmosphère-Climat (IMT Atlantique Rennes), February 6th. Panel of the Machine Learning workshop at CRiP ITES, Deauville, April 5th.

Cécile Germain - Evaluator for the H2020-2016-CNECT program; member of the DFG review panel within Germany's excellence strategy selection process.

Isabelle Guyon - Member of the NeurIPS foundation board.

Marc Schoenauer - Comité Scientifique IA, SCube (Scientipôle Savoirs & Société), Orsay; Scientific Committee, TrackML (see Section 7.6); Comité de sélection, Chaire ABEONA-ENS "Biais et Equité en IA"; Conseil Scientifique, IFPEN; Scientific Advisory Board, BCAM, Bilbao, Spain; Scientific Advisory Board, Tara Oceans, Paris.

Michèle Sebag - Hiring juries : LRI; Centrale-Supélec; ENS-Paris; UCA-Nice; U. Freiburg, Germany. Selection juries: Awards NIPS 2019. Propositions NSERC, Canada; Propositions Dpt STIC; Prix de thèse AFIA. Expert Committee from Finland's Minister of Economic Affairs, AI Strategy, February 2019.

10.1.7. Research Administration

Isabelle Guyon - Representative of UPSud in the DataIA *Institut de Convergence* Program Committee, University of Paris-Saclay. Responsible of Master AIC (becoming Paris-Saclay master in Artificial Intelligence).

Marc Schoenauer - Deputy Scientific Director of Inria (in French, Directeur Scientifique Adjoint, DSA), in charge of AI.

Michele Sebag - Deputy director of LRI, CNRS UMR 8623; elected member of the Research Council of Univ. Paris-Saclay; member of the STIC department council of Univ. Paris-Saclay; member of the Scientific Council of Labex AMIES, Applications des Mathématiques ds l'Industrie, l'Entreprise et la Société.

Paola Tubaro - Representative of CNRS in the DataIA *Institut de Convergence* Program Committee, University of Paris-Saclay; member of the Board, Maison des Sciences de l'Homme Paris-Saclay; member of CLIP, Institut Pascal, University of Paris-Saclay.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : Philippe Caillou, Computer Science for students in Accounting and Management, 192h, L1, IUT Sceaux, Univ. Paris Sud.

Licence : Aurélien Decelle, Computer Architecture, 42h, L2, Univ. Paris-Sud.

Licence : Aurélien Decelle, Introduction to Machine Learning, 57h, L2, Univ. Paris-Sud.

Licence : François Landes, Mathematics for Computer Scientists, 51h, L2, Univ. Paris-Sud.

Licence : François Landes, Intro to Machine Learning, 48h, L2, Univ. Paris-Sud.

Licence and Polytech : Cécile Germain, Computer Architecture

Licence : Isabelle Guyon: Introduction to Data Science, L1, Univ. Paris-Sud.

Licence : Isabelle Guyon, Project: Resolution of mini-challenges (created by M2 students), L2, Univ. Paris-Sud.

Master : François Landes, Machine Learning, 34h, M1 Polytech, U. Paris-sud.

Master : Aurélien Decelle, Machine Learning, 26h, M1, Univ. Paris-Sud.

Master : Aurélien Decelle, Probability and statistics, 26h, M1, Univ. Paris-Sud.

Master : Aurélien Decelle, Probability, statistics and information theory, M1, Univ. Paris-Sud.

Master : Guillaume Charpiat, Deep Learning in Practice, 24h, M2 Recherche, Centrale-Supelec + MVA.

Master : Isabelle Guyon, Project: Creation of mini-challenges, M2, Univ. Paris-Sud.

Master : Michèle Sebag, Machine Learning, 12h; Deep Learning, 9h; Reinforcement Learning, 12h; M2 Recherche, U. Paris-Sud.

Master : François Landes, Machine Learning, 22h, M2 Recherche, U. Paris-Sud.

Master : Paola Tubaro, Start -up project for engineering students, 24h, Telecom ParisTech.

Master : Paola Tubaro, Sociology of social networks, 24h, M2, EHESS/ENS.

Master : Paola Tubaro, Social and economic network science, 24h, M2, ENSAE.

Doctorate: Paola Tubaro, Research Methods, 12h, University of Insubria, Italy.

Summer school: Guillaume Charpiat, Machine Learning & Deep Learning Tutorial, 4h30, ForMaL, ENS Cachan, June 4th

10.2.2. Supervision

HdR - Paola Tubaro, *Decoding the platform society: Organizations, markets and networks in the digital economy*, 11/12/2019, Sciences Po Paris.

PhD - Benjamin DONNOT, *Deep learning methods for predicting flows in power grids : novel architectures and algorithms.*, 13/02/2019, Isabelle Guyon and Antoine Marot (RTE)

PhD - Corentin TALLEC, *Reinforcement Learning and Recurrent Neural Networks: Dynamical approaches*, 7/10/2019, Université Paris-Saclay, Yann Ollivier

PhD - Mandar CHANDORKAR, *Machine Learning in Space Weather*, 14/11/2019, University of Eindhoven, Enrico Camporeale, Cyril Furtlehner, and Michèle Sebag

PhD - Guillaume DOQUET, *Agnostic Feature Selection*, 29/11/2019, Université Paris-Saclay, Michèle Sebag

PhD - Diviyan KALAINATHAN, *Generative Neural Networks to Infer Causal Mechanisms: Algorithms and Applications*, 17/12/2019, Université Paris-Saclay, Michèle Sebag and Isabelle Guyon

PhD - Lisheng SUN, *Meta-Learning as a Markov Decision Process*, 19/12/2019, Université Paris-Saclay, Isabelle Guyon and Michèle Sebag

PhD in progress - Eléonore BARTENLIAN, *Deep Learning pour le traitement du signal*, 1/10/2018, Michèle Sebag and Frédéric Pascal (Centrale-Supélec)

PhD in progress - Victor BERGER, *Variational Anytime Simulator*, 1/10/2017, Michèle Sebag

PhD in progress - Guillaume BIED, *Valorisation des Données pour la Recherche d'Emploi*, 1/10/2019, Bruno Crepon (CREST-ENSAE) and Philippe Caillou

PhD in progress - Leonard BLIER, *Vers une architecture stable pour les systèmes d'apprentissage par renforcement*, 1/09/2018, Yann Ollivier (Facebook AI Research, Paris) and Marc Schoenauer

PhD in progress - Tony BONNAIRE, *Reconstruction de la toile cosmique*, from 1/10/2018, Nabila Aghanim (Institut d'Astrophysique Spatiale) and Aurélien Decelle

PhD in progress - Balthazar DONON, *Apprentissage par renforcement pour une conduite stratégique du système électrique*, 1/10/2018, Isabelle Guyon and Antoine Marot (RTE)

PhD in progress - Victor ESTRADE *Robust domain-adversarial learning, with applications to High Energy Physics,* 01/10/2016, Cécile Germain and Isabelle Guyon.

PhD in progress - Loris FELARDOS, *Neural networks for molecular dynamics simulations*, 1/10/2018, Guillaume Charpiat, Jérôme Hénin (IBPC) and Bruno Raffin (InriAlpes)

PhD in progress - Giancarlo FISSORE, *Statistical physics analysis of generative models*, 1/10/2017, Aurélien Decelle and Cyril Furtlehner

PhD in progress - Julien GIRARD, *Vérification et validation des techniques d'apprentissage automatique*, 1/10/2018, Zakarian Chihani (CEA) and Guillaume Charpiat

PhD in progress - Nicolas GIRARD, *Satellite image vectorization using neural networks*, 1/10/2017, Yuliya Tarabalka & Pierre Alliez (Inria Sophia-Antipolis) and Guillaume Charpiat

PhD in progress - Armand LACOMBE, *Recommandation de Formations: Application de l'apprentissage causal dans le domaine des ressources humaines*, 1/10/2019, Michele Sebag and Philippe Caillou

PhD in progress - Zhengying LIU, *Automation du design des reseaux de neurones profonds*, 1/10/2017, Isabelle Guyon

PhD in progress - Nizam MAKDOUD, *Motivations intrinsèques en apprentissage par renforcement. Application à la recherche de failles de sécurité*, 1/02/2018, Marc Schoenauer and Jérôme Kodjabachian (ThalèsThereSIS, Palaiseau).

PhD in progress - Marc NABHAN, *Sûreté de fonctionnement d'un véhicule autonome - évaluation des fausses détections au travers d'un profil de mission réduit*, 1/10/2017, Marc Schoenauer and Hiba Hage (Renault)

PhD in progress - Adrian POL *Machine Learning Anomaly Detection, with application to CMS Data Quality Monitoring,* 01/10/2016, Cécile Germain.

PhD in progress - Herilalaina RAKOTOARISON, *Automatic Algorithm Configuration for Power Grid Optimization*, 1/10/2017, Marc Schoenauer and Michèle Sebag

PhD in progress - Théophile SANCHEZ, *Reconstructing the past: deep learning for population genetics*, 1/10/2017, Guillaume Charpiat and Flora Jay

PhD in progress - Pierre WOLINSKI, *Learning the Architecture of Neural Networks*, from 1/9/2016, Yann Ollivier (Facebook AI Research, Paris) and Guillaume Charpiat

PhD in progress - Wenzhuo LIU, Machine Learning for Numerical Simulation of PDEs, from 1/11/2019, Mouadh Yagoubi (IRT SystemX) and Marc Schoenauer

PhD in progress - Marion ULLMO, *Reconstruction de la toile cosmique*, from 1/10/2018, Nabila Aghanim (Institut d'Astrophysique Spatiale) and Aurélien Decelle

10.2.3. Juries

Marc Schoenauer: Reviewer for Dennis Wilson, IRIT, Université Toulouse; PhD jury of Patricio Cerda Reyes, U. Paris-Saclay, 28/11/2019.

François Landes: PhD rapporteur of Martina Teruzzi (Condensed matter and Machine Learning PhD, at SISSA, Trieste, Italy).

Paola Tubaro: PhD jury of Sophie Balech, U. Paris Nanterre, 09/07/2019; PhD jury of Linda Rua, U. Paris Dauphine, 13/12/2019.

Guillaume Charpiat: half-way PhD committee ("à mi-parcours") of Rodrigo Daudt (ONERA), of Julie Rivet (EPITA) and of Nissim Zerbib (Institut de la Vision)

Isabelle Guyon: PhD jury Justine Falque, U. Paris-Saclay (29/11/2019).

Michele Sebag: Reviewer for A. Chemla (PhD U. Roma & IRCAM); Pierre Fournier (PhD ISIR).

10.3. Popularization

10.3.1. Articles and contents

- Michèle Sebag - *Il faut dissiper le malentendu sur "les prétentions infondées" de l'intelligence artificielle*, tribune du journal *Le Monde* du 12 août 2019.
- Paola Tubaro - BBC World service, *The 'microworkers' making your digital life possible*, 02/08/2019; Radio France Inter, *Travailleurs du clic, les soutiers du clavier*, 23/03/2019; Science & Vie, *La force de l'IA repose sur des travailleurs bien r'eels*, 24/10/2019; AFP/L'Express, *Les travailleurs du clic, petites mains invisibles de l'économie numérique*, 21/06/2019; CNRS Le Journal, *Ces microtravailleurs de l'ombre*, 24/05/2019; Le Monde, *Jobs du clic : la France compte plus de 250 000 micro-travailleurs*, 15/02/2019; L'Humanité Dimanche, *Travailleuse du clic, la double invisibilité*, 02/05/2019.

10.3.2. Interventions

- Guillaume Charpiat - *Génomes présents, histoires d'antan: Apprentissage profond pour la génétique des populations*, Inria Seminar Unithé ou Café with Flora Jay, Feb 2019.
- Marc Schoenauer - *Intelligence Artificielle Mythes et réalités*, Classe virtuelle de la Délégation Académique du Numérique Educatif, Feb. 2, 2019; *Intelligence Artificielle Mythes et réalités*, Médiathèque George Sand, Palaiseau, Feb. 9, 2019.
- Michèle Sebag - Journées Scientifiques Inria, Lyon 2019; Centre de Recherches Interdisciplinaires, Workshop on Artificial Intelligence and Women Empowerment (Oct. 2019); CRI, *Open AI: From big data to smart data* (Nov. 2019).

- Paola Tubaro - *Le Micro-Travail en France : derrière l'automatisation, de nouvelles précarités au travail ?* CFE-CGC Orange, Paris, 01/07/2019; *Le Micro-Travail en France*, Fondation Gabriel Péri, Paris, 26/06/2019; *Big data et société : vie privée, travail, inégalités*, Banque de France, 06/06/2019.

11. Bibliography

Major publications by the team in recent years

[1] C. ADAM-BOURDARIOS, G. COWAN, C. GERMAIN-RENAUD, I. GUYON, B. KÉGL, D. ROUSSEAU.*The Higgs Machine Learning Challenge*, in "Journal of Physics: Conference Series", December 2015, vol. 664, n^{o} 7 [*DOI :* 10.1088/1742-6596/664/7/072015], https://hal.inria.fr/hal-01745998

[2] L. DA COSTA, Á. FIALHO, M. SCHOENAUER, M. SEBAG.*Adaptive Operator Selection with Dynamic Multi-Armed Bandits*, in "Proc. Genetic and Evolutionary Computation Conference (GECCO)", ACM, 2008, p. 913-920, ACM-SIGEVO 10-years Impact Award [*DOI :* 10.1145/1389095.1389272], https://hal.inria.fr/inria-00278542

[3] C. FURTLEHNER, A. DECELLE.*Cycle-based Cluster Variational Method for Direct and Inverse Inference*, in "Journal of Statistical Physics", August 2016, vol. 164, n^{o} 3, p. 531–574, https://hal.inria.fr/hal-01214155

[4] S. GELLY, M. SCHOENAUER, M. SEBAG, O. TEYTAUD, L. KOCSIS, D. SILVER, C. SZEPESVARI.*The Grand Challenge of Computer Go: Monte Carlo Tree Search and Extensions*, in "Communications- ACM", 2012, vol. 55, n^{o} 3, p. 106-113, https://hal.inria.fr/hal-00695370

[5] O. GOUDET, D. KALAINATHAN, P. CAILLOU, D. LOPEZ-PAZ, I. GUYON, M. SEBAG.*Learning Functional Causal Models with Generative Neural Networks*, in "Explainable and Interpretable Models in Computer Vision and Machine Learning", Springer Series on Challenges in Machine Learning, Springer International Publishing, 2018, https://arxiv.org/abs/1709.05321 [*DOI :* 10.1007/978-3-319-98131-4], https://hal.archives-ouvertes.fr/hal-01649153

[6] T. LUCAS, C. TALLEC, J. VERBEEK, Y. OLLIVIER.*Mixed batches and symmetric discriminators for GAN training*, in "ICML - 35th International Conference on Machine Learning", Stockholm, Sweden, July 2018, https://hal.inria.fr/hal-01791126

[7] E. MAGGIORI, Y. TARABALKA, G. CHARPIAT, P. ALLIEZ.*Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification*, in "IEEE Transactions on Geoscience and Remote Sensing", 2017, vol. 55, n^{o} 2, p. 645-657, https://hal.inria.fr/hal-01369906

[8] M. MISIR, M. SEBAG.*Alors: An algorithm recommender system*, in "Artificial Intelligence", 2017, vol. 244, p. 291-314, Published on-line Dec. 2016, https://hal.inria.fr/hal-01419874

[9] Y. OLLIVIER, L. ARNOLD, A. AUGER, N. HANSEN.*Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles*, in "Journal of Machine Learning Research", 2017, vol. 18, n^{o} 18, p. 1-65, https://hal.inria.fr/hal-01515898

[10] X. ZHANG, C. FURTLEHNER, C. GERMAIN-RENAUD, M. SEBAG.*Data Stream Clustering with Affinity Propagation*, in "IEEE Transactions on Knowledge and Data Engineering", 2014, vol. 26, n^{o} 7, 1, https://hal.inria.fr/hal-00862941

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. CHANDORKAR.*Machine Learning in Space Weather*, Université of Eindhoven, November 2019, https://hal.inria.fr/tel-02430788

[12] B. DONNOT.*Deep learning methods for predicting flows in power grids : novel architectures and algorithms*, Université Paris-Saclay, February 2019, https://tel.archives-ouvertes.fr/tel-02045873

[13] G. F. DOQUET.*Agnostic Feature Selection*, Université Paris-Saclay/Université Paris-Sud, November 2019, https://hal.archives-ouvertes.fr/tel-02436845

[14] D. KALAINATHAN.*Generative Neural networks to infer Causal Mechanisms: Algorithms and applications*, Université Paris Sud (Paris 11) - Université Paris Saclay, December 2019, https://hal.inria.fr/tel-02435986

[15] L. SUN-HOSOYA.*Meta-Learning as a Markov Decision Process*, Université Paris-Saclay, December 2019, https://hal.archives-ouvertes.fr/tel-02422144

[16] C. TALLEC.*Recurrent Neural Networks and Reinforcement Learning: Dynamic Approaches*, Université Paris-Saclay, October 2019, https://hal.inria.fr/tel-02434367

Articles in International Peer-Reviewed Journal

[17] B. DONON, B. DONNOT, I. GUYON, Z. LIU, A. MAROT, P. PANCIATICI, M. SCHOENAUER.*LEAP Nets for System Identification and Application to Power Systems*, in "Neurocomputing", 2019, To appear, forthcoming, https://hal.inria.fr/hal-02422708

[18] S. GIFFARD-ROISIN, M. YANG, G. CHARPIAT, C. K. BONFANTI, B. KÉGL, C. MONTELEONI.*Tropical Cyclone Track Forecasting using Fused Deep Learning from Aligned Reanalysis Data*, in "Frontiers in Big Data", January 2020, https://arxiv.org/abs/1910.10566 , forthcoming, https://hal.archives-ouvertes.fr/hal-02329437

[19] F. P. LANDES, G. BIROLI, O. DAUCHOT, A. LIU, D. REICHMAN.*Attractive versus truncated repulsive supercooled liquids: The dynamics is encoded in the pair correlation function*, in "Physical Review E ", January 2020, vol. 101, n^{o} 1, https://arxiv.org/abs/1906.01103 [*DOI :* 10.1103/PHYSREVE.101.010602], https://hal.inria.fr/hal-02439295

[20] E. LIPPIELLO, G. PETRILLO, F. P. LANDES, A. ROSSO.*Fault Heterogeneity and the Connection between Aftershocks and Afterslip*, in "Bulletin of the Seismological Society of America", April 2019, vol. 109, n^{o} 3, p. 1156-1163, https://arxiv.org/abs/1812.05862 [*DOI :* 10.1785/0120180244], https://hal.inria.fr/hal-02156407

[21] A. A. POL, G. CERMINARA, C. GERMAIN, M. PIERINI, A. SETH.*Detector Monitoring with Artificial Neural Networks at the CMS Experiment at the CERN Large Hadron Collider*, in "Computing and Software for Big Science", January 2019, vol. 3, n^{o} 1 [*DOI :* 10.1007/S41781-018-0020-1], https://hal.inria.fr/hal-02427998

[22] P. TUBARO, A. A. CASILLI.*Micro-work, artificial intelligence and the automotive industry*, in "Journal of Industrial and Business Economics = Economia e politica industriale", June 2019, p. 1-13 [*DOI :* 10.1007/S40812-019-00121-1], https://hal.archives-ouvertes.fr/hal-02148979

[23] P. TUBARO.*La vie privée, un bien commun ?*, in "Regards croisés sur l'économie", 2019, vol. 23, n° 2, p. 129-137 [*DOI :* 10.3917/RCE.023.0129], https://hal.archives-ouvertes.fr/hal-02196333

[24] P. TUBARO.*Les Learning Analytics vus par la sociologie*, in "Distances et Médiations des Savoirs", December 2019, n° 28, https://hal.archives-ouvertes.fr/hal-02418562

[25] P. TUBARO.*Whose results are these anyway? Reciprocity and the ethics of "giving back" after social network research*, in "Social Networks", November 2019 [*DOI :* 10.1016/J.SOCNET.2019.10.003], https://hal.archives-ouvertes.fr/hal-02360709

International Conferences with Proceedings

[26] V. BERGER, M. SEBAG.*From abstract items to latent spaces to observed data and back: Compositional Variational Auto-Encoder*, in "ECML PKDD 2019 - European Conference on Machine learning and knowledge discovery in databases", Würzburg, Germany, September 2019, https://arxiv.org/abs/2001.07910 , https://hal.inria.fr/hal-02431955

[27] L. BLIER, P. WOLINSKI, Y. OLLIVIER.*Learning with Random Learning Rates*, in "ECML PKDD 2019 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Würzburg, Germany, September 2019, https://arxiv.org/abs/1810.01322 - 16 pages, 4 figures, code available on GitHub, https://hal.archives-ouvertes.fr/hal-01888352

[28] M. CHANDORKAR, C. FURTLEHNER, B. PODUVAL, E. CAMPOREALE, M. SEBAG.*Dynamic Time Lag Regression: Predicting What and When*, in "ICLR 2020 - 8th International Conference on Learning Representations", Addis Abeba, Ethiopia, April 2020, https://hal.inria.fr/hal-02422148

[29] G. CHARPIAT, N. GIRARD, L. FELARDOS, Y. TARABALKA.*Input Similarity from the Neural Network Perspective*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, https://hal.inria.fr/hal-02394647

[30] B. DONNOT, B. DONON, I. GUYON, Z. LIU, A. MAROT, P. PANCIATICI, M. SCHOENAUER.*LEAP nets for power grid perturbations*, in "ESANN 2019 - 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning", Bruges, Belgium, April 2019, https://arxiv.org/abs/1908.08314 , https://hal.archives-ouvertes.fr/hal-02268886

[31] B. DONON, B. DONNOT, I. GUYON, A. MAROT.*Graph Neural Solver for Power Systems*, in "IJCNN 2019 - International Joint Conference on Neural Networks", Budapest, Hungary, July 2019, https://hal.archives-ouvertes.fr/hal-02175989

[32] *Best Paper*
G. F. DOQUET, M. SEBAG.*Agnostic feature selection*, in "ECML PKDD 2019 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Würzburg, Germany, September 2019, https://hal.archives-ouvertes.fr/hal-02436824.

[33] N. GIRARD, G. CHARPIAT, Y. TARABALKA.*Noisy Supervision for Correcting Misaligned Cadaster Maps Without Perfect Ground Truth Data*, in "IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium", Yokohama, Japan, July 2019, https://arxiv.org/abs/1903.06529 , https://hal.inria.fr/hal-02065211

[34] J. GIRARD-SATABIN, G. CHARPIAT, Z. CHIHANI, M. SCHOENAUER.*CAMUS: A Framework to Build Formal Specifications for Deep Perception Systems Using Simulators*, in "ECAI 2020 - 24th European Conference on Artificial Intelligence", Santiago de Compostela, Spain, June 2020, https://hal.inria.fr/hal-02440520

[35] A. KHALEL, O. TASAR, G. CHARPIAT, Y. TARABALKA.*Multi-Task Deep Learning for Satellite Image Pansharpening and Segmentation*, in "IGARSS 2019 - IEEE International Geoscience and Remote Sensing Symposium", Yokohama, Japan, July 2019, https://hal.inria.fr/hal-02276549

[36] A. MAROT, A. ROSIN, L. CROCHEPIERRE, B. DONNOT, P. PINSON, L. BOUDJELOUD-ASSALA.*Interpreting atypical conditions in systems with deep conditional Autoencoders: the case of electrical consumption*, in "ECML PKDD 2019 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Wurzburg, Germany, September 2019, https://hal.archives-ouvertes.fr/hal-02266963

[37] M. NABHAN, M. SCHOENAUER, Y. TOURBIER, H. HAGE.*Optimizing coverage of simulated driving scenarios for the autonomous vehicle*, in "IEEE ICCVE 2019 - International Conference on Connected Vehicles and Expo", Graz, Austria, November 2019, https://hal.inria.fr/hal-02433530

[38] A. PAVAO, D. KALAINATHAN, L. SUN-HOSOYA, K. P. BENNETT, I. GUYON.*Design and Analysis of Experiments: A Challenge Approach in Teaching*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, https://hal.inria.fr/hal-02415639

[39] A. A. POL, V. BERGER, G. CERMINARA, C. GERMAIN, M. PIERINI.*Anomaly Detection With Conditional Variational Autoencoders*, in "ICMLA 2019 - 18th IEEE International Conference on Machine Learning and Applications", Boca Raton, United States, 18th International Conference on Machine Learning Applications, December 2019, https://hal.inria.fr/hal-02396279

[40] H. RAKOTOARISON, M. SCHOENAUER, M. SEBAG.*Automated Machine Learning with Monte-Carlo Tree Search*, in "IJCAI-19 - 28th International Joint Conference on Artificial Intelligence", Macau, China, International Joint Conferences on Artificial Intelligence Organization, August 2019, p. 3296-3303 [*DOI :* 10.24963/IJCAI.2019/457], https://hal.inria.fr/hal-02300884

[41] A. SCHOENAUER SEBAG, L. HEINRICH, M. SCHOENAUER, M. SEBAG, L. WU, S. ALTSCHULER.*Multi-Domain Adversarial Learning*, in "ICLR 2019 - Seventh annual International Conference on Learning Representations", New Orleans, United States, Tara Sainath, May 2019, https://hal.inria.fr/hal-01968180

[42] C. TALLEC, L. BLIER, Y. OLLIVIER.*Making Deep Q-learning methods robust to time discretization*, in "ICML 2019 - Thirty-sixth International Conference on Machine Learning", Long Beach, United States, PMLR, 2019, n^{o} 97, p. 6096-6104, https://arxiv.org/abs/1901.09732 , https://hal.inria.fr/hal-02435523

[43] P. TUBARO.*Emergent relational structures at a "sharing economy" festival*, in "COMPLEX NETWORKS 2018 - The 7th International Conference on Complex Networks and Their Applications", Cambridge, United Kingdom, Springer, 2019, vol. 2, p. 559-571, https://hal.archives-ouvertes.fr/hal-01947911

[44] A. YALE, S. DASH, R. DUTTA, I. GUYON, A. PAVAO, K. P. BENNETT.*Privacy Preserving Synthetic Health Data*, in "ESANN 2019 - European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning", Bruges, Belgium, April 2019, https://hal.inria.fr/hal-02160496

Conferences without Proceedings

[45] Z. LIU, I. GUYON, J. C. S. JACQUES JUNIOR, M. MADADI, S. ESCALERA, A. PAVAO, H. J. ESCALANTE, W.-W. TU, Z. XU, S. TREGUER.*AutoCV Challenge Design and Baseline Results*, in "CAp 2019 - Conférence sur l'Apprentissage Automatique", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02265053

Scientific Books (or Scientific Book chapters)

[46] S. AMROUCHE, L. BASARA, P. CALAFIURA, V. ESTRADE, S. FARRELL, D. FERREIRA, L. FINNIE, N. FINNIE, C. GERMAIN, V. V. GLIGOROV, T. GOLLING, S. GORBUNOV, H. GRAY, I. GUYON, M. HUSHCHYN, V. INNOCENTE, M. KIEHN, E. MOYSE, J.-F. PUGET, Y. REINA, D. ROUSSEAU, A. SALZBURGER, A. USTYUZHANIN, J.-R. VLIMANT, J. S. WIND, T. XYLOURIS, Y. YILMAZ.*The Tracking Machine Learning Challenge: Accuracy Phase*, in "The NeurIPS '18 Competition", The Springer Series on Challenges in Machine Learning, Springer, November 2019, vol. 8, p. 231-264 [*DOI :* 10.1007/978-3-030-29135-8_9], https://hal.inria.fr/hal-02427989

[47] O. GOUDET, D. KALAINATHAN, M. SEBAG, I. GUYON.*Learning Bivariate Functional Causal Models*, in "Cause Effect Pairs in Machine Learning", I. GUYON, A. STATNIKOV, B. B. BATU (editors), The Springer Series on Challenges in Machine Learning, Springer Verlag, October 2019, p. 101-153 [*DOI :* 10.1007/978-3-030-21810-2_3], https://hal.inria.fr/hal-02433201

[48] I. GUYON, A. STATNIKOV, B. B. BATU.*Cause Effect Pairs in Machine Learning*, The Springer Series on Challenges in Machine Learning, Springer Verlag, 2019 [*DOI :* 10.1007/978-3-030-21810-2], https://hal.inria.fr/hal-02433195

[49] I. GUYON, O. GOUDET, D. KALAINATHAN.*Evaluation Methods of Cause-Effect Pairs*, in "Cause Effect Pairs in Machine Learning", I. GUYON, A. STATNIKOV, B. B. BATU (editors), The Springer Series on Challenges in Machine Learning, Springer Verlag, October 2019, p. 27-99 [*DOI :* 10.1007/978-3-030-21810-2_2], https://hal.inria.fr/hal-02433198

[50] I. GUYON, A. STATNIKOV.*Results of the Cause-Effect Pair Challenge*, in "Cause Effect Pairs in Machine Learning", I. GUYON, A. STATNIKOV, B. B. BATU (editors), The Springer Series on Challenges in Machine Learning. Springer, Springer Verlag, October 2019, p. 237-256 [*DOI :* 10.1007/978-3-030-21810-2_7], https://hal.inria.fr/hal-02433204

[51] D. KALAINATHAN, O. GOUDET, M. SEBAG, I. GUYON.*Discriminant Learning Machines*, in "Cause Effect Pairs in Machine Learning", I. GUYON, A. STATNIKOV, B. B. BATU (editors), The Springer Series on Challenges in Machine Learning, Springer Verlag, October 2019, p. 155-189 [*DOI :* 10.1007/978-3-030-21810-2_4], https://hal.inria.fr/hal-02433203

Research Reports

[52] A. A. CASILLI, P. TUBARO, C. LE LUDEC, M. COVILLE, M. BESENVAL, T. MOUHTARE, E. WAHAL.*Micro-work in France. Behind Automation, New Forms of Precarious Labour?*, Projet de recherche DiPLab, May 2019, https://hal.archives-ouvertes.fr/hal-02139528

Other Publications

[53] T. Bonnaire, N. Aghanim, A. Decelle, M. Douspis. *T-ReX: a graph-based filament detection method*, December 2019, https://arxiv.org/abs/1912.00732 - Submitted to A&A, 16 pages, 16 figures, https://hal.archives-ouvertes.fr/hal-02420696

[54] A. Decelle, C. Furtlehner. *Gaussian-Spherical Restricted Boltzmann Machines*, December 2019, https://arxiv.org/abs/1910.14544 - 27 pages, 8 figures, https://hal.archives-ouvertes.fr/hal-02420813

[55] A. Decelle, S. Hwang, J. Rocchi, D. Tantari. *Inverse problems for structured datasets using parallel TAP equations and RBM*, December 2019, https://arxiv.org/abs/1906.11988 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02420755

[56] B. Donon, R. Clément, B. Donnot, A. Marot, I. Guyon, M. Schoenauer. *Neural Networks for Power Flow : Graph Neural Solver*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02372741

[57] H. J. Escalante, H. Kaya, A. A. Salah, S. Escalera, Y. Gucluturk, U. Güçlü, X. Baró, I. Guyon, J. C. S. Jacques Junior, M. Madadi, S. Ayache, E. Viegas, F. Gurpinar, A. S. Wicaksana, C. C. S. Liem, M. A. J. van Gerven, R. van Lier. *Explaining First Impressions: Modeling, Recognizing, and Explaining Apparent Personality from Videos*, January 2019, https://arxiv.org/abs/1802.00745 - Accepted at Transaction on Affective Computing, https://hal.inria.fr/hal-01991652

[58] G. Fissore, A. Decelle, C. Furtlehner, Y. Han. *Robust Multi-Output Learning with Highly Incomplete Data via Restricted Boltzmann Machines*, December 2019, https://arxiv.org/abs/1912.09382 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02420824

[59] O. François, S. Liégeois, B. Demaille, F. Jay. *Inference of population genetic structure from temporal samples of DNA*, December 2019, biorxiv [*DOI :* 10.1101/801324], https://hal.archives-ouvertes.fr/hal-02413974

[60] J. Girard-Satabin, G. Charpiat, Z. Chihani, M. Schoenauer. *CAMUS: A Framework to Build Formal Specifications for Deep Perception Systems Using Simulators*, November 2019, https://arxiv.org/abs/1911.10735 - working paper or preprint, https://hal.inria.fr/hal-02374956

[61] C. Le Ludec, P. Tubaro, A. A. Casilli. *Combien de personnes micro-travaillent en France ? Estimer l'ampleur d'une nouvelle forme de travail*, February 2019, Il s'agit de la version française du document arxiv 1901.03889 (HAL 02012731). Publié dans les documents de travail i3, https://hal.archives-ouvertes.fr/hal-02021525

[62] C. Le Ludec, P. Tubaro, A. A. Casilli. *How many people microwork in France? Estimating the size of a new labor force*, January 2019, https://arxiv.org/abs/1901.03889 - 9 pages, 1 figure, 7 tables. The article is part of the DiPLab ("Digital Platform Labor") research project. For more information: http://diplab.eu/Il s'agit de la version anglaise du document hal-02021525, https://hal.archives-ouvertes.fr/hal-02012731

[63] Z. Liu, Z. Xu, S. Escalera, I. Guyon, J. C. S. Jacques Junior, M. Madadi, A. Pavao, S. Treguer, W.-W. Tu. *Towards Automated Computer Vision: Analysis of the AutoCV Challenges 2019*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02386805

[64] A. A. POL, V. BERGER, G. CERMINARA, C. GERMAIN, M. PIERINI.*Trigger Rate Anomaly Detection with Conditional Variational Autoencoders at the CMS Experiment*, December 2019, Machine Learning and the Physical Sciences Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS), Poster, https://hal.inria.fr/hal-02428005

[65] A. A. POL, G. CERMINARA, C. GERMAIN, M. PIERINI, A. SETH.*Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider*, January 2019, https://arxiv.org/abs/1808.00911 - working paper or preprint, https://hal.inria.fr/hal-01976256

[66] B. YELMEN, A. DECELLE, L. ONGARO, D. MARNETTO, F. MONTINARO, C. FURTLEHNER, L. PAGANI, F. JAY.*Creating Artificial Human Genomes Using Generative Models*, December 2019, biorxiv [*DOI :* 10.1101/769091], https://hal.archives-ouvertes.fr/hal-02413942

References in notes

[67] C. ADAM-BOURDARIOS, G. COWAN, C. GERMAIN, I. GUYON, B. KÉGL, D. ROUSSEAU.*How Machine Learning won the Higgs Boson Challenge*, in "Proc. European Symposium on ANN, CI and ML", 2016

[68] C. ADAM-BOURDARIOS, G. COWAN, C. GERMAIN-RENAUD, I. GUYON, B. KÉGL, D. ROUSSEAU.*The Higgs Machine Learning Challenge*, in "Journal of Physics: Conference Series", 2015, vol. 664, n^{o} 7 [*DOI :* 10.1088/1742-6596/664/7/072015], https://hal.inria.fr/hal-01745998

[69] D. J. AMIT, H. GUTFREUND, H. SOMPOLINSKY.*Statistical Mechanics of Neural Networks Near Saturation*, in "Annals of Physics", 1987, vol. 173, p. 30-67

[70] C. ANDERSON.*The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, 2008, Wired Magazine, https://www.wired.com/2008/06/pb-theory/

[71] N. BELKHIR.*Per Instance Algorithm Configuration for Continuous Black Box Optimization*, Université Paris-Saclay, November 2017

[72] N. BELKHIR, J. DRÉO, P. SAVÉANT, M. SCHOENAUER.*Per instance algorithm configuration of CMA-ES with limited budget*, in "Proc. ACM-GECCO", 2017, p. 681-688

[73] I. BELLO, B. ZOPH, V. VASUDEVAN, Q. V. LE.*Neural Optimizer Search with Reinforcement Learning*, in "34th ICML", 2017, p. 459–468

[74] S. BEN-DAVID, J. BLITZER, K. CRAMMER, A. KULESZA, F. PEREIRA, J. W. VAUGHAN.*A theory of learning from different domains*, in "Machine Learning", 2018, vol. 79, n^{o} 1, p. 151–175 [*DOI :* 10.1007/s10994-009-5152-4]

[75] J. BERGSTRA, R. BARDENET, Y. BENGIO, B. KÉGL.*Algorithms for Hyper-Parameter Optimization*, in "NIPS 25", 2011, p. 2546–2554

[76] J. BRUNA, S. MALLAT.*Invariant Scattering Convolution Networks*, in "IEEE Trans. Pattern Anal. Mach. Intell.", 2013, vol. 35, n^{o} 8, p. 1872–1886

[77] J.-J. CHRISTOPHE, J. DECOCK, O. TEYTAUD.*Direct model predictive control*, in "European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)", 2014, https://hal.inria.fr/hal-00958192

[78] R. COUILLET, M. DEBBAH.*Random matrix methods for wireless communications*, Cambridge University Press, 2011

[79] DAVID ROUSSEAU ET AL..*The TrackML challenge*, in "NIPS 2018 - 32nd Annual Conference on Neural Information Processing Systems", Montreal, Canada, December 2018, p. 1-23, https://hal.inria.fr/hal-01745714

[80] DAVID ROUSSEAU ET AL..*The TrackML Particle Tracking Challenge*, July 2018, https://hal.inria.fr/hal-01680537

[81] DAVID SILVER ET AL..*Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm*, in "CoRR", 2017, vol. abs/1712.01815, http://arxiv.org/abs/1712.01815

[82] M. DEMARIA, M. MAINELLI, L. K. SHAY, J. A. KNAFF, J. KAPLAN.*Further improvements to the statistical hurricane intensity prediction scheme (SHIPS)*, in "Weather and Forecasting", 2005, vol. 20, n^{o} 4, p. 531–543

[83] A. DECELLE, G. FISSORE, C. FURTLEHNER.*Spectral dynamics of learning in restricted Boltzmann machines*, in "EPL (Europhysics Letters)", 2017, vol. 119, n^{o} 6, 60001

[84] A. DECELLE, G. FISSORE, C. FURTLEHNER.*Thermodynamics of Restricted Boltzmann Machines and Related Learning Dynamics*, in "J. Stat. Phys.", 2018, vol. 172, p. 1576-1608

[85] L. DINH, J. SOHL-DICKSTEIN, S. BENGIO.*Density estimation using Real NVP*, in "Int. Conf. on Learning Representations (ICLR)", 2017

[86] B. DONNOT, I. GUYON, M. SCHOENAUER, A. MAROT, P. PANCIATICI.*Anticipating contingengies in power grids using fast neural net screening*, in "IEEE International Joint Conference on Neural Networks", Rio de Janeiro, Brazil, Proc. 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, July 2018, https://arxiv.org/abs/1805.02608 , https://hal.archives-ouvertes.fr/hal-01783669

[87] B. DONNOT, I. GUYON, M. SCHOENAUER, A. MAROT, P. PANCIATICI.*Fast Power system security analysis with Guided Dropout*, in "26th European Symposium on Artificial Neural Networks", Bruges, Belgium, Electronic Proceedings ESANN 2018, April 2018, https://arxiv.org/abs/1801.09870 , https://hal.archives-ouvertes.fr/hal-01695793

[88] T. DURIEZ, S. L. BRUNTON, B. R. NOACK.*Machine Learning Control – Taming Nonlinear Dynamics and Turbulence*, Springer International Publishing, 2017

[89] D. K. DUVENAUD, D. MACLAURIN, J. AGUILERA-IPARRAGUIRRE, R. GÓMEZ-BOMBARELLI, T. HIRZEL, A. ASPURU-GUZIK, R. P. ADAMS.*Convolutional Networks on Graphs for Learning Molecular Fingerprints*, in "NIPS", 2015, p. 2224–2232

[90] T. V. ERVEN, P. GRUNWALD, S. DE ROOIJ.*Catching up faster by switching sooner: a predictive approach to adaptive estimation with an application to the AIC-BIC dilemma: Catching up Faster*, in "J. Royal Statistical Society: B", 2012, vol. 74, n^{o} 3, p. 361–417 [*DOI :* 10.1111/J.1467-9868.2011.01025.X]

[91] H. J. ESCALANTE, I. GUYON, S. ESCALERA, J. JACQUES, M. MADADI, X. BARÓ, S. AYACHE, E. VIEGAS, Y. GÜÇLÜTÜRK, U. GÜÇLÜ, M. A. J. VAN GERVEN, R. VAN LIER.*Design of an Explainable Machine Learning Challenge for Video Interviews*, in "IJCNN 2017 - 30th International Joint Conference on Neural Networks", Anchorage, AK, United States, Neural Networks (IJCNN), 2017 International Joint Conference on, IEEE, 2017, p. 1-8, https://hal.inria.fr/hal-01668386

[92] S. ESCALERA, X. BARÓ, H. J. ESCALANTE, I. GUYON.*ChaLearn looking at people: A review of events and resources*, in "2017 International Joint Conference on Neural Networks (IJCNN)", 2017, p. 1594-1601

[93] V. ESTRADE, C. GERMAIN, I. GUYON, D. ROUSSEAU.*Systematics aware learning: a case study in High Energy Physics*, in "ESANN 2018 - 26th European Symposium on Artificial Neural Networks", Bruges, Belgium, April 2018, https://hal.inria.fr/hal-01715155

[94] S. FALKNER, A. KLEIN, F. HUTTER.*BOHB: Robust and Efficient Hyperparameter Optimization at Scale*, in "Proc. 35th ICML", 2018, p. 1436–1445

[95] M. FEURER, A. KLEIN, K. EGGENSPERGER, J. SPRINGENBERG, M. BLUM, F. HUTTER.*Efficient and Robust Automated Machine Learning*, in "NIPS 28", 2015, p. 2962–2970

[96] L. FRANCESCHI, P. FRASCONI, S. SALZO, R. GRAZZI, M. PONTIL.*Bilevel Programming for Hyperparameter Optimization and Meta-Learning*, in "Proc. 35th ICML", J. G. DY, A. KRAUSE (editors), 2018, p. 1563–1572

[97] C. FURTLEHNER, A. DECELLE.*Cycle-Based Cluster Variational Method for Direct and Inverse Inference*, in "Journal of Statistical Physics", 2016, vol. 164, n^{o} 3, p. 531–574

[98] C. FURTLEHNER, M. SEBAG, X. ZHANG.*Scaling analysis of affinity propagation*, in "Physical Review E", 2010, vol. 81, n^{o} 6, 066102

[99] Y. GANIN, E. USTINOVA, H. AJAKAN, P. GERMAIN, H. LAROCHELLE, F. LAVIOLETTE, M. MARCHAND, V. LEMPITSKY.*Domain-Adversarial Training of Neural Networks*, in "Journal of Machine Learning Research", 2016, vol. 17, n^{o} 59, p. 1-35

[100] K. GARIMELLA, G. D. F. MORALES, A. GIONIS, M. MATHIOUDAKIS.*Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship*, in "WWW", ACM, 2018, p. 913–922

[101] S. GIFFARD-ROISIN, M. YANG, G. CHARPIAT, B. KÉGL, C. MONTELEONI.*Deep Learning for Hurricane Track Forecasting from Aligned Spatio-temporal Climate Datasets*, in "Modeling and decision-making in the spatiotemporal domain NIPS workhop", Montréal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01905408

[102] J. GILES.*Computational social science: Making the links*, in "Nature - News", 2012, vol. 488, n^{o} 7412, p. 448-450

[103] F. GONARD.*Cold-start recommendation : from Algorithm Portfolios to Job Applicant Matching*, Université Paris-Saclay, May 2018, https://tel.archives-ouvertes.fr/tel-01825220

[104] F. GONARD, M. SCHOENAUER, M. SEBAG.*ASAP.V2 and ASAP.V3: Sequential optimization of an Algorithm Selector and a Scheduler*, in "Open Algorithm Selection Challenge 2017 ", Proceedings of Machine Learning Research, 2017, vol. 79, p. 8-11

[105] I. GOODFELLOW, J. POUGET-ABADIE, M. MIRZA, B. XU, D. WARDE-FARLEY, S. OZAIR, A. COURVILLE, Y. BENGIO.*Generative Adversarial Nets*, in "NIPS 27", 2014, p. 2672–2680

[106] I. GOODFELLOW, J. POUGET-ABADIE, M. MIRZA, B. XU, D. WARDE-FARLEY, S. OZAIR, A. COURVILLE, Y. BENGIO.*Generative Adversarial Nets*, in "NIPS 27", Z. GHAHRAMANI, M. WELLING, C. CORTES, N. D. LAWRENCE, K. Q. WEINBERGER (editors), Curran Associates, Inc., 2014, p. 2672–2680

[107] O. GOUDET, D. KALAINATHAN, P. CAILLOU, D. LOPEZ-PAZ, I. GUYON, M. SEBAG.*Learning Functional Causal Models with Generative Neural Networks*, in "Explainable and Interpretable Models in Computer Vision and Machine Learning", Springer Series on Challenges in Machine Learning, Springer International Publishing, 2018 [*DOI :* 10.1007/978-3-319-98131-4], https://hal.archives-ouvertes.fr/hal-01649153

[108] P. GRUNWALD.*The Minimum Description Length Principle*, MIT Press, 2007

[109] I. GUYON, C. F. ALIFERIS, G. F. COOPER, A. ELISSEEFF, J.-P. PELLET, P. SPIRTES, A. R. STATNIKOV.*Design and Analysis of the Causation and Prediction Challenge*, in "WCCI Causation and Prediction Challenge", JMLR W-CP, 2008, p. 1–33

[110] I. GUYON, K. BENNETT, G. CAWLEY, H. J. ESCALANTE, S. ESCALERA, T. K. HO, N. MACIA, B. RAY, M. SAEED, A. STATNIKOV.*Design of the 2015 ChaLearn AutoML challenge*, in "Proc. IJCNN", IEEE, 2015, p. 1–8

[111] I. GUYON, K. BENNETT, G. CAWLEY, H. J. ESCALANTE, S. ESCALERA, T. K. HO, N. MACIA, B. RAY, M. SAEED, A. STATNIKOV.*Design of the 2015 ChaLearn AutoML challenge*, in "Proc. IJCNN", IEEE, 2015, p. 1–8

[112] I. GUYON, L. BOTTOU, B. SCHÖLKOPF, A. STATNIKOV, E. VIEGAS, J. M. ROBINS.*NIPS workshop on causality*, 2013, http://clopinet.com/isabelle/Projects/NIPS2013/

[113] I. GUYON, L. SUN-HOSOYA, M. BOULLÉ, H. J. ESCALANTE, S. ESCALERA, Z. LIU, D. JAJETIC, B. RAY, M. SAEED, M. SEBAG, A. STATNIKOV, W.-W. TU, E. VIEGAS.*Analysis of the AutoML Challenge series 2015-2018*, in "AutoML: Methods, Systems, Challenges", F. HUTTER, L. KOTTHOFF, J. VANSCHORE (editors), The Springer Series on Challenges in Machine Learning, Springer Verlag, 2018, forthcoming, https://hal.archives-ouvertes.fr/hal-01906197

[114] D. HERNÁNDEZ-LOBATO, P. MORALES-MOMBIELA, D. LOPEZ-PAZ, A. SUÁREZ.*Non-linear Causal Inference using Gaussianity Measures*, in "Journal of Machine Learning Research", 2016, vol. 17, n^{o} 28, p. 1-39, http://jmlr.org/papers/v17/14-375.html

[115] H. H. HOOS.*Programming by Optimization*, in "Commun. ACM", 2012, vol. 55, n^{o} 2, p. 70–80

[116] D. HOTHERSALL.*History of Psychology*, McGraw-Hill, 2004

[117] G. W. IMBENS, D. B. RUBIN.*Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015

[118] A. JACOT, F. GABRIEL, C. HONGLER.*Neural Tangent Kernel: Convergence and Generalization in Neural Networks*, in "Advances in Neural Information Processing Systems 31", 2018, p. 8571–8580

[119] D. KALAINATHAN, O. GOUDET, P. CAILLOU, M. SEBAG, P. TUBARO.*Conditions objectives de travail et ressenti des individus : le rôle du management*, 2017, 12, Synthèse n. 14 de ”La Fabrique de l’Industrie”, https://hal.inria.fr/hal-01742592

[120] D. KALAINATHAN, O. GOUDET, I. GUYON, D. LOPEZ-PAZ, M. SEBAG.*SAM: Structural Agnostic Model, Causal Discovery and Penalized Adversarial Learning*, August 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01864239

[121] D. KINGMA, M. WELLING.*Auto-encoding variational Bayes.*, in "Int. Conf. on Learning Representations (ICLR)", 2014

[122] L. KOTTHOFF.*Algorithm Selection for Combinatorial Search Problems: A Survey*, in "Data Mining and Constraint Programming: Foundations of a Cross-Disciplinary Approach", CH. BESSIERE ET AL. (editor), Springer Verlag, 2016, p. 149–190

[123] A. KRIZHEVSKY, I. SUTSKEVER, G. E. HINTON.*ImageNet Classification with Deep Convolutional Neural Networks*, in "Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1", NIPS’12, 2012, p. 1097–1105

[124] D. LAZER, A. PENTLAND, L. ADAMIC, S. ARAL, A. L. BARABASI, D. BREWER, N. CHRISTAKIS, N. CONTRACTOR, J. FOWLER, M. GUTMANN, T. JEBARA, G. KING, M. MACY, D. ROY, M. VAN ALSTYNE.*Life in the network: the coming age of computational social science*, in "Science", 2009, vol. 323, n^{o} 5915, p. 721–723

[125] Y. LECUN.*Deep Learning est mort. Vive Differentiable Programming!*, 2018, Facebook Post, https://www.facebook.com/yann.lecun/posts/10155003011462143

[126] J. Z. LIANG, E. MEYERSON, R. MIIKKULAINEN.*Evolutionary architecture search for deep multitask networks*, in "Proc. ACM-GECCO", H. E. AGUIRRE, K. TAKADAMA (editors), ACM, 2018, p. 466–473, http://doi.acm.org/10.1145/3205455.3205489

[127] T. P. LILLICRAP, J. J. HUNT, A. PRITZEL, N. HEESS, T. EREZ, Y. TASSA, D. SILVER, D. WIERSTRA.*Continuous control with deep reinforcement learning*, 2015

[128] H. LIPSON, J. B. POLLACK.*Automatic design and manufacture of robotic lifeforms*, in "Nature Letters", 2000, vol. 406, p. 974–978

[129] Z. LIU, O. BOUSQUET, A. ELISSEEFF, S. ESCALERA, I. GUYON, J. JACQUES, A. PAVAO, D. SILVER, L. SUN-HOSOYA, S. TREGUER, W.-W. TU, J. WANG, Q. YAO.*AutoDL Challenge Design and Beta Tests-Towards automatic deep learning*, in "CiML workshop @ NIPS2018", Montreal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01906226

[130] D. LOPEZ-PAZ, K. MUANDET, B. SCHÖLKOPF, I. O. TOLSTIKHIN.*Towards a Learning Theory of Cause-Effect Inference*, in "Proceedings of the 32nd International Conference on Machine Learning, ICML", F. R. BACH, D. M. BLEI (editors), 2015, p. 1452–1461, http://jmlr.org/proceedings/papers/v37/lopez-paz15.html

[131] C. LOUIZOS, M. REISSER, T. BLANKEVOORT, E. GAVVES, M. WELLING.*Relaxed Quantization for Discretized Neural Networks*, in "ICLR", 2019

[132] T. LUCAS, C. TALLEC, J. VERBEEK, Y. OLLIVIER.*Mixed batches and symmetric discriminators for GAN training*, in "ICML - 35th International Conference on Machine Learning", Stockholm, Sweden, July 2018, https://hal.inria.fr/hal-01791126

[133] E. MAGGIORI, Y. TARABALKA, G. CHARPIAT, P. ALLIEZ.*Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark*, in "IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)", Fort Worth, United States, July 2017, https://hal.inria.fr/hal-01468452

[134] A. MAROT, B. DONON, I. GUYON, B. DONNOT.*Learning To Run a Power Network Competition*, in "CiML Workshop, NeurIPS", Montréal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01968295

[135] L. MESCHEDER, A. GEIGER, S. NOWOZIN.*Which Training Methods for GANs do actually Converge?*, in "35th ICML", 2018, p. 3481–3490

[136] D. MOLCHANOV, A. ASHUKHA, D. VETROV.*Variational Dropout Sparsifies Deep Neural Networks*, in "ArXiv e-prints", January 2017

[137] M. MISIR, M. SEBAG.*Alors: An algorithm recommender system*, in "Artificial Intelligence", 2017, vol. 244, p. 291-314, Published on-line Dec. 2016, https://hal.inria.fr/hal-01419874

[138] C. O'NEILL.*Weapons of Math Destruction*, Crown Books, 2016

[139] J. PEARL.*Causality: Models, Reasoning, and Inference (2nd edition)*, Cambridge University Press, 2009

[140] J. PEARL.*Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*, in "11th ACM WSDM", 2018, 3

[141] G. PLEISS, M. RAGHAVAN, F. WU, J. M. KLEINBERG, K. Q. WEINBERGER.*On Fairness and Calibration*, in "NIPS", 2017, p. 5684–5693

[142] M. RAISSI.*Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations*, in "JMLR", 2018, vol. 19, p. 1–24

[143] G. RAMÍREZ, E. VILLATORO, B. IONESCU, H. J. ESCALANTE, S. ESCALERA, M. LARSON, H. MÜLLER, I. GUYON.*Overview of the Multimedia Information Processing for Personality and Social Networks Analysis Contest*, in "International Conference on Pattern Recognition (ICPR)", T. TAN, J. KITTLER, A. JAI (editors), IEEE, 2018, p. 127-139

[144] S. RAVANBAKHSH, J. SCHNEIDER, B. PÓCZOS.*Equivariance Through Parameter-Sharing*, in "International Conference on Machine Learning", 2017, p. 2892–2901, http://proceedings.mlr.press/v70/ravanbakhsh17a.html

[145] E. REAL, S. MOORE, A. SELLE, S. SAXENA, Y. L. SUEMATSU, J. T. Q. LE, A. KURAKIN.*Large-Scale Evolution of Image Classifiers*, in "34th ICML", 2017

[146] J. R. RICE.*The Algorithm Selection Problem*, in "Advances in Computers", 1976, vol. 15, p. 65 – 118 [*DOI :* 10.1016/S0065-2458(08)60520-3], http://www.sciencedirect.com/science/article/pii/S0065245808605203

[147] J. RISSANEN.*Information and Complexity in Statistical Modeling*, Information Science and Statistics, Springer-Verlag, 2007

[148] T. SANCHEZ, G. CHARPIAT, F. JAY.*SPI-DNA: End-to-end Deep Learning Approach for Demographic History Inference*, in "Paris-Saclay Junior Conference on Data Science and Engineering", Orsay, France, September 2017, https://hal.archives-ouvertes.fr/hal-01679385

[149] M. SCHMIDT, H. LIPSON.*Distilling free-form natural laws from experimental data*, in "Science", 2009, vol. 324, n^{o} 5923, p. 81–85

[150] S. SIGTIA, E. BENETOS, S. DIXON.*An End-to-End Neural Network for Polyphonic Piano Music Transcription*, in "IEEE/ACM Trans. Audio, Speech & Language Processing", 2016, vol. 24, n^{o} 5, p. 927–939

[151] J. SIRIGNANO, K. SPILIOPOULOS.*DGM: A deep learning algorithm for solving partial differential equations*, in "Journal of Computational Physics", 2018, vol. 375, p. 1339 – 1364 [*DOI :* 10.1016/J.JCP.2018.08.029]

[152] I. STENGERS.*Another Science Is Possible*, Open Humanities Press, 2013

[153] D. H. STERN, R. HERBRICH, T. GRAEPEL.*Matchbox: large scale online bayesian recommendations*, in "WWW", ACM Press, 2010, 111 [*DOI :* 10.1145/1526709.1526725]

[154] L. SUN-HOSOYA, I. GUYON, M. SEBAG.*Lessons learned from the AutoML challenge*, in "Conférence sur l'Apprentissage Automatique 2018", Rouen, France, June 2018, https://hal.inria.fr/hal-01811454

[155] C. SUNSTEIN, R. THALER.*Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, 2008

[156] C. SZEGEDY, W. ZAREMBA, I. SUTSKEVER, J. BRUNA, D. ERHAN, I. J. GOODFELLOW, R. FERGUS.*Intriguing properties of neural networks*, in "CoRR", 2013, vol. abs/1312.6199, http://arxiv.org/abs/1312.6199

[157] C. TALLEC, Y. OLLIVIER.*Can recurrent neural networks warp time?*, in "ICLR", 2018, https://openreview.net/forum?id=SJcKhk-Ab

[158] C. TALLEC, Y. OLLIVIER.*Can recurrent neural networks warp time?*, in "International Conference on Learning Representation 2018", Vancouver, France, April 2018, https://hal.inria.fr/hal-01812064

[159] C. TALLEC, Y. OLLIVIER.*Unbiased online recurrent optimization*, in "International Conference On Learning Representation", Vancouver, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01972587

[160] S. TRIANTAFILLOU, I. TSAMARDINOS.*Constraint-based Causal Discovery from Multiple Interventions over Overlapping Variable Sets*, in "Journal of Machine Learning Research", 2015, vol. 16, p. 2147-2205

[161] J. VANSCHOREN, J. N. VAN RIJN, B. BISCHL, L. TORGO.*OpenML: networked science in machine learning*, in "SIGKDD Explorations", 2013, vol. 15, n^o 2, p. 49–60, https://arxiv.org/abs/1407.7722

[162] VOLODYMYR MNIH ET AL..*Human-level control through deep reinforcement learning*, in "Nature", 2015, vol. 518, n^o 7540, p. 529–533, https://doi.org/10.1038/nature14236

[163] J. WAN, S. ESCALERA, G. ANBARJAFARI, H. J. ESCALANTE, X. BARÓ, I. GUYON, M. MADADI, J. ALLIK, C. LIN, Y. XIE.*Results and Analysis of ChaLearn LAP Multi-modal Isolated and Continuous Gesture Recognition, and Real versus Fake Expressed Emotions Challenges*, in "International Conference on Computer Vision - ICCV 2017", 2017, https://hal.inria.fr/hal-01677974

[164] M. WEIMER, A. KARATZOGLOU, Q. LE, A. SMOLA.*CoFiRank - Maximum Margin Matrix Factorization for Collaborative Ranking*, in "21th NIPS", 2007, p. 222-230

[165] M. WELLING.*Intelligence per Kilowatthour*, 2018, Keynote talk @ICML, https://icml.cc/Conferences/2018/Schedule?showEvent=1866

[166] F. M. J. WILLEMS, Y. M. SHTARKOV, T. J. TJALKENS.*The context-tree weighting method: basic properties*, in "IEEE Transactions on Information Theory", 1995, vol. 41, n^o 3, p. 653-664

[167] A. ZAMPIERI, G. CHARPIAT, N. GIRARD, Y. TARABALKA.*Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing*, in "European Conference on Computer Vision (ECCV)", Munich, Germany, September 2018, https://hal.inria.fr/hal-01849389

[168] X. ZHANG, C. FURTLEHNER, C. GERMAIN-RENAUD, M. SEBAG.*Data Stream Clustering with Affinity Propagation*, in "IEEE Transactions on Knowledge and Data Engineering", 2014, vol. 26, n^o 7, 1, https://hal.inria.fr/hal-00862941

[169] M. DE BEZENAC, A. PAJOT, P. GALLINARI.*Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge*, in "arXiv preprint arXiv:1711.07970", 2017

[170] A. VAN DEN OORD, S. DIELEMAN, H. ZEN, K. SIMONYAN, O. VINYALS, A. GRAVES, N. KALCHBRENNER, A. SENIOR, K. KAVUKCUOGLU.*WaveNet: A Generative Model for Raw Audio*, in "CoRR", 2016, vol. abs/1609.03499, http://arxiv.org/abs/1609.03499

Project-Team TOCCATA

Formally Verified Programs, Certified Tools, Numerical Computations

IN COLLABORATION WITH: Laboratoire de recherche en informatique (LRI)

IN PARTNERSHIP WITH:
CNRS

Université Paris-Sud (Paris 11)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Proofs and Verification

Table of contents

Project-Team TOCCATA

Creation of the Team: 2012 September 01, updated into Project-Team: 2014 July 01

The Toccata team (http://toccata.lri.fr/) is a research team common to Inria Saclay–Île-de-France, CNRS, and Université Paris-Sud. Team members are also members of the larger VALS research group (Verification of Algorithms, Languages and systems; http://vals.lri.fr/) of the LRI (Laboratoire de Recherche en Informatique, UMR 8623).

Keywords:

Computer Science and Digital Science:

A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.6. - Concurrent programming
A2.1.10. - Domain-specific languages
A2.1.11. - Proof languages
A2.4.2. - Model-checking
A2.4.3. - Proofs
A6.2.1. - Numerical analysis of PDE and ODE
A7.2. - Logic in Computer Science
A7.2.1. - Decision procedures
A7.2.2. - Automated Theorem Proving
A7.2.3. - Interactive Theorem Proving
A7.2.4. - Mechanized Formalization of Mathematics
A8.10. - Computer arithmetic

Other Research Topics and Application Domains:

B5.2.2. - Railway
B5.2.3. - Aviation
B5.2.4. - Aerospace
B6.1. - Software industry
B9.5.1. - Computer science
B9.5.2. - Mathematics

1. Team, Visitors, External Collaborators

Research Scientists

Claude Marché [Team leader, Inria, Senior Researcher, HDR]
Sylvie Boldo [Inria, Senior Researcher, HDR]
Jean-Christophe Filliâtre [CNRS, Senior Researcher, HDR]
Guillaume Melquiond [Inria, Researcher, HDR]

Faculty Members

Sylvain Conchon [Univ Paris-Sud, Professor, HDR]
Thibault Hilaire [Délégation de l'Univ Pierre et Marie Curie, Associate Professor, until Aug 2019]
Andrei Paskevich [Univ Paris-Sud, Associate Professor]

External Collaborators

Thibaut Balabonski [Univ Paris-Sud, Associate Professor]
Chantal Keller [Univ Paris-Sud, Associate Professor]
Benjamin Farinier [Univ Grenoble Alpes, A.T.E.R, until Sep 2019]

Technical Staff

Benedikt Becker [Inria, Engineer]
Cláudio Lourenço [Inria, Engineer, from Sep 2019]
Sylvain Dailler [Inria, Engineer]
Rehan Malak [Inria, Engineer, from Apr 2019 until Sep 2019]

PhD Students

Albin Coquereau [École Nationale Supérieure de Techniques Avancées, until Jun 2019]
Florian Faissole [Inria, until Sep 2019]
Diane Gallois-Wong [Univ Paris-Sud]
Quentin Garchery [Univ Paris-Sud]
Antoine Lanco [Inria, from Oct 2019]
Mário Pereira [Grant Portuguese government, until Apr 2019]
Raphaël Rieu-Helft [CIFRE TrustInSoft]
Mattias Roux [Univ Paris-Sud]

Post-Doctoral Fellows

Cláudio Lourenço [Univ Paris-Sud, until Aug 2019]
Florian Steinberg [Inria, until Sep 2019]

Administrative Assistant

Katia Evrat

2. Overall Objectives

2.1. Presentation

The general objective of the Toccata project is to promote formal specification and computer-assisted proof in the development of software that requires high assurance in terms of safety and correctness with respect to its intended behavior. Such safety-critical software appears in many application domains like transportation (e.g., aviation, aerospace, railway, and more and more in cars), communication (e.g., internet, smartphones), health devices, etc. The number of tasks performed by software is quickly increasing, together with the number of lines of code involved. Given the need of high assurance of safety in the functional behavior of such applications, the need for automated (i.e., computer-assisted) methods and techniques to bring guarantee of safety became a major challenge. In the past and at present, the most widely used approach to check safety of software is to apply heavy test campaigns, which take a large part of the costs of software development. Yet they cannot ensure that all the bugs are caught, and remaining bugs may have catastrophic causes (e.g., the Heartbleed bug in OpenSSL library discovered in 2014 https://en.wikipedia.org/wiki/Heartbleed).

Generally speaking, software verification approaches pursue three goals: (1) verification should be sound, in the sense that no bugs should be missed, (2) verification should not produce false alarms, or as few as possible, (3) it should be as automatic as possible. Reaching all three goals at the same time is a challenge. A large class of approaches emphasizes goals (2) and (3): testing, run-time verification, symbolic execution, model checking, etc. Static analysis, such as abstract interpretation, emphasizes goals (1) and (3). Deductive verification emphasizes (1) and (2). The Toccata project is mainly interested in exploring the deductive verification approach, although we also consider the other ones in some cases.

In the past decade, there have been significant progress made in the domain of deductive program verification. They are emphasized by some success stories of application of these techniques on industrial-scale software. For example, the *Atelier B* system was used to develop part of the embedded software of the Paris metro line 14 [50] and other railway-related systems; a formally proved C compiler was developed using the Coq proof assistant [63]; the L4-verified project developed a formally verified micro-kernel with high security guarantees, using analysis tools on top of the Isabelle/HOL proof assistant [62]. A bug in the JDK implementation of TimSort was discovered using the KeY environment [69] and a fixed version was proved sound. Another sign of recent progress is the emergence of deductive verification competitions (e.g., VerifyThis [1]). Finally, recent trends in the industrial practice for development of critical software is to require more and more guarantees of safety, e.g., the new DO-178C standard for developing avionics software adds to the former DO-178B the use of formal models and formal methods. It also emphasizes the need for certification of the analysis tools involved in the process.

3. Research Program

3.1. Research Program

3.1.1. Panorama of Deductive Verification

There are two main families of approaches for deductive verification. Methods in the first family build on top of mathematical proof assistants (e.g., Coq, Isabelle) in which both the model and the program are encoded; the proof that the program meets its specification is typically conducted in an interactive way using the underlying proof construction engine. Methods from the second family proceed by the design of standalone tools taking as input a program in a particular programming language (e.g., C, Java) specified with a dedicated annotation language (e.g., ACSL [49], JML [56]) and automatically producing a set of mathematical formulas (the *verification conditions*) which are typically proved using automatic provers (e.g., Z3 [70], Alt-Ergo [58], CVC4 [48]).

The first family of approaches usually offers a higher level of assurance than the second, but also demands more work to perform the proofs (because of their interactive nature) and makes them less easy to adopt by industry. Moreover, they generally do not allow to directly analyze a program written in a mainstream programming language like Java or C. The second kind of approaches has benefited in the past years from the tremendous progress made in SAT and SMT solving techniques, allowing more impact on industrial practices, but suffers from a lower level of trust: in all parts of the proof chain (the model of the input programming language, the VC generator, the back-end automatic prover), potential errors may appear, compromising the guarantee offered. Moreover, while these approaches are applied to mainstream languages, they usually support only a subset of their features.

3.1.2. Overall Goals of the Toccata Project

One of our original skills is the ability to conduct proofs by using automatic provers and proof assistants at the same time, depending on the difficulty of the program, and specifically the difficulty of each particular verification condition. We thus believe that we are in a good position to propose a bridge between the two families of approaches of deductive verification presented above. Establishing this bridge is one of the goals of the Toccata project: we want to provide methods and tools for deductive program verification that can offer both a high amount of proof automation and a high guarantee of validity. Indeed, an axis of research of Toccata is the development of languages, methods and tools that are themselves formally proved correct.

In industrial applications, numerical calculations are very common (e.g. control software in transportation). Typically they involve floating-point numbers. Some of the members of Toccata have an internationally recognized expertise on deductive program verification involving floating-point computations. Our past work includes a new approach for proving behavioral properties of numerical C programs using Frama-C/Jessie [47], various examples of applications of that approach [54], the use of the Gappa solver for proving numerical algorithms [68], an approach to take architectures and compilers into account when dealing with floating-point programs [55], [66]. We also contributed to the Handbook of Floating-Point Arithmetic [65]. A representative case study is the analysis and the proof of both the method error and the rounding error of a numerical analysis program solving the one-dimension acoustic wave equation [52] [51]. Our experience led us to a conclusion that verification of numerical programs can benefit a lot from combining automatic and interactive theorem proving [53], [54], [59]. Verification of numerical programs is another main axis of Toccata.

Our scientific programme detailed below is structured into four axes:

1. Foundations and spreading of deductive program verification;
2. Reasoning on mutable memory in program verification;
3. Verification of Computer Arithmetic;
4. Spreading Formal Proofs.

Let us conclude with more general considerations about our agenda of the next four years: we want to keep on

- with general audience actions;
- industrial transfer, in particular through an extension of the perimeter of the ProofInUse joint lab.

3.2. Foundations and spreading of deductive program verification

Permanent researchers: S. Conchon, J.-C. Filliâtre, C. Marché, G. Melquiond, A. Paskevich

This axis covers the central theme of the team: deductive verification, from the point of view of its foundations but also our will to spread its use in software development. The general motto we want to defend is "deductive verification for the masses". A non-exhaustive list of subjects we want to address is as follows.

- The verification of general-purpose algorithms and data structures: the challenge is to discover adequate invariants to obtain a proof, in the most automatic way as possible, in the continuation of the current VOCaL project and the various case studies presented in Axis 4 below.
- Uniform approaches to obtain correct-by-construction programs and libraries, in particular by automatic extraction of executable code (in OCaml, C, CakeML, etc.) from verified programs, and including innovative general methods like advanced ghost code, ghost monitoring, etc.
- Automated reasoning dedicated to deductive verification, so as to improve proof automation; improved combination of interactive provers and fully automated ones, proof by reflection.
- Improved feedback in case of proof failures: based on generation of counterexamples, or symbolic execution, or possibly randomized techniques à la quickcheck.
- Reduction of the trusted computing base in our toolchains: production of certificates from automatic proofs, for goal transformations (like those done by Why3), and from the generation of VCs

A significant part of the work achieved in this axis is related to the Why3 toolbox and its ecosystem, displayed on Figure 1. The boxes in red background correspond to the tools we develop in the Toccata team.

3.3. Reasoning on mutable memory in program verification

Permanent researchers: J.-C. Filliâtre, C. Marché, G. Melquiond, A. Paskevich

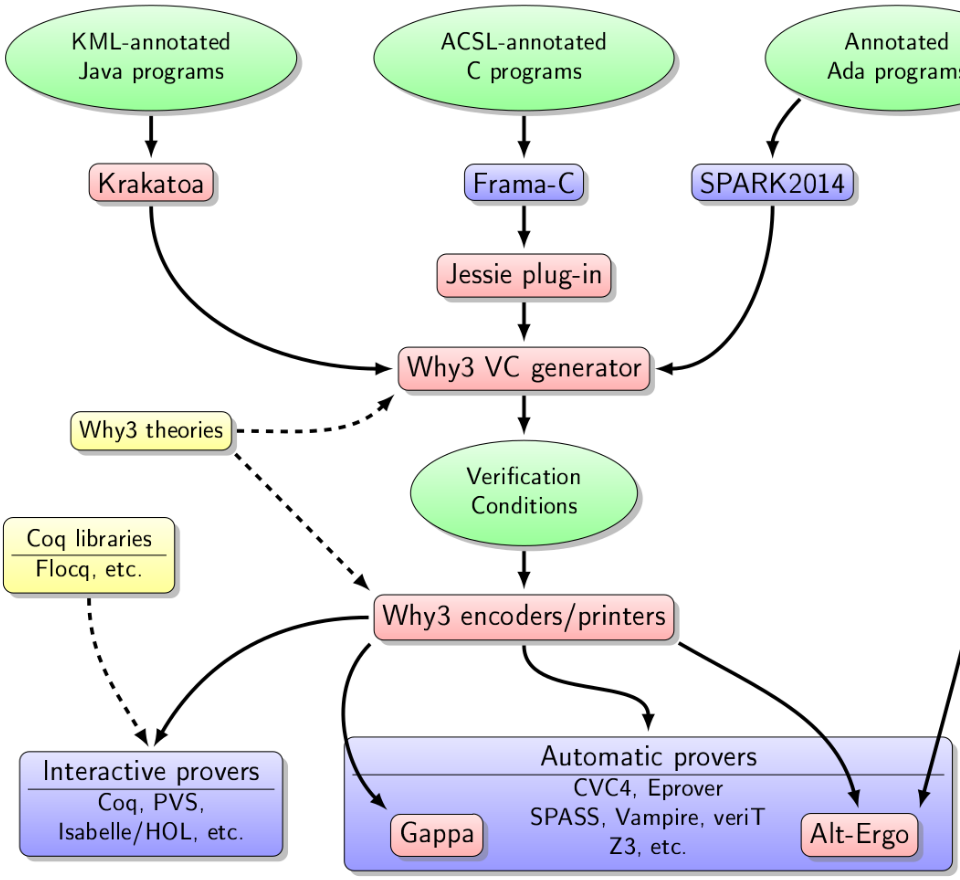


Figure 1. The Why3 ecosystem

This axis concerns specifically the techniques for reasoning on programs where aliasing is the central issue. It covers the methods based on type-based alias analysis and related memory models, on specific program logics such as separation logics, and extended model-checking. It concerns the application on analysis of C or C++ codes, on Ada codes involving pointers, but also concurrent programs in general. The main topics planned are:

- The study of advanced type systems dedicated to verification, for controlling aliasing, and their use for obtaining easier-to-prove verification conditions. Modern typing system in the style of Rust, involving ownership and borrowing, will be considered.
- The design of front-ends of Why3 for the proofs of programs where aliasing cannot be fully controlled statically, via adequate memory models, aiming in particular at extraction to C; and also for concurrent programs.
- The continuation of fruitful work on concurrent parameterized systems, and its corresponding specific SMT-based model-checking.
- Concurrent programming on weak memory models, on one hand as an extension of parameterized systems above, but also in the specific context of OCaml multicore (http://ocamllabs.io/doc/multicore.html).
- In particular in the context of the ProofInUse joint lab, design methods for Ada, C, C++ or Java using memory models involving fine-grain analysis of pointers. Rust programs could be considered as well.

3.4. Verification of Computer Arithmetic

Permanent researchers: S. Boldo, C. Marché, G. Melquiond

We of course want to keep this axis which is a major originality of Toccata. The main topics of the next 4 years will be:

- Fundamental studies concerning formalization of floating-point computations, algorithms, and error analysis. Related to numerical integration, we will develop the relationships between mathematical stability and floating-point stability of numerical schemes.
- A significant effort dedicated to verification of numerical programs written in C, Ada, C++. This involves combining specifications in real numbers and computation in floating-point, and underlying automated reasoning techniques with floating-point numbers and real numbers. A new approach we have in mind concerns some variant of symbolic execution of both code and specifications involving real numbers.
- We have not yet studied embedded systems. Our approach is first to tackle numerical filters. This requires more results on fixed-point arithmetic and a careful study of overflows.
- Also a specific focus on arbitrary precision integer arithmetic, in the continuation of the ongoing PhD thesis of R. Rieu-Helft.

3.5. Spreading Formal Proofs

Permanent researchers: S. Boldo, S. Conchon, J.-C. Filliâtre, C. Marché, G. Melquiond, A. Paskevich

This axis covers applications in general. The applications we currently have in mind are:

- Hybrid Systems, i.e., systems mixing discrete and continuous transitions. This theme covers many aspects such as general techniques for formally reasoning of differential equations, and extending SMT-based reasoning. The challenge is to support both abstract mathematical reasoning and concrete program execution (e.g., using floating-point representation). Hybrid systems will be a common effort with other members of the future laboratory joint with LSV of ENS Cachan.

- Applied mathematics, in the continuation of the current efforts towards verification of Finite Element Method. It has only been studied in the mathematical point of view during this period. We plan to also consider their floating-point behavior and a demanding application is that of molecular simulation exhibited in the new EMC2 project. The challenge here is both in the mathematics to be formalized, in the numerical errors that have never been studied (and that may be huge in specific cases), and in the size of the programs, which requires that our tools scale.
- Continuation of our work on analysis of shell scripts. The challenge is to be able to analyze a large number of scripts (more than 30,000 in the corpus of Debian packages installation scripts) in an automatic manner. An approach that will be considered is some form of symbolic execution.
- Explore proof tools for mathematics, in particular automated reasoning for real analysis (application: formalization of the weak Goldbach conjecture), and in number theory.
- Obtain and distribute verified OCaml libraries, as expected outcome of the VOCaL project.
- Formalization of abstract interpretation and WP calculi: in the continuation of the former project Verasco, and an ongoing project proposal joint with CEA List. The difficulty of achieving full verification of such tools will be mitigated by use of certificate techniques.

4. Application Domains

4.1. Safety-Critical Software

The application domains we target involve safety-critical software, that is where a high-level guarantee of soundness of functional execution of the software is wanted. Currently our industrial collaborations or impact mainly belong to the domain of transportation: aerospace, aviation, railway, automotive.

Transfer to aeronautics: Airbus France Development of the control software of Airbus planes historically includes advanced usage of formal methods. A first aspect is the usage of the CompCert verified compiler for compiling C source code. Our work in cooperation with Gallium team for the safe compilation of floating-point arithmetic operations [2] is directly in application in this context. A second aspect is the usage of the Frama-C environment for static analysis to verify the C source code. In this context, both our tools Why3 and Alt-Ergo are indirectly used to verify C code.

Transfer to the community of Atelier B In the former ANR project BWare, we investigated the use of Why3 and Alt-Ergo as an alternative back-end for checking proof obligations generated by *Atelier B*, whose main applications are railroad-related https://www.atelierb.eu/en/. The transfer effort continues nowadays through the FUI project LCHIP.

ProofInUse joint lab: transfer to the community of Ada development Through the creation of the ProofInUse joint lab (https://www.adacore.com/proofinuse) in 2014, with AdaCore company (https://www.adacore.com/), we have a growing impact on the community of industrial development of safety-critical applications written in Ada. See the web page https://www.adacore.com/industries for a an overview of AdaCore's customer projects, in particular those involving the use of the SPARK Pro tool set. This impact involves both the use of Why3 for generating VCs on Ada source codes, and the use of Alt-Ergo for performing proofs of those VCs.

The impact of ProofInUse can also be measured in term of job creation: the first two ProofInUse engineers, D. Hauzar and C. Fumex, employed initially on Inria temporary positions, have now been hired on permanent positions in AdaCore company. It is also interesting to notice that this effort allowed AdaCore company to get new customers, in particular the domains of application of deductive formal verification went beyond the historical domain of aerospace: application in automotive (https://www.adacore.com/customers/toyota-itc-japan) cyber-security (https://www.adacore.com/customers/multi-level-security-workstation), health (artificial heart, https://www.adacore.com/customers/total-artificial-heart).

Extension of ProofInUse joint lab The current plans for continuation of the ProofInUse joint lab (https://why3.gitlabpages.inria.fr/proofinuse/) include extension at a larger perimeter than Ada applications. We started to collaborate with the TrustInSoft company (https://trust-in-soft.com/) for the verification of C and C++ codes, including the use of Why3 to design verified and reusable C libraries (ongoing CIFRE PhD thesis). We also started to collaborate with Mitsubishi Electric in Rennes (http://www.mitsubishielectric-rce.eu/xindex.php) for a specific usage of Why3 for verifying embedded devices (logic controllers).

Generally speaking, we believe that our increasing industrial impact is a representative success for our general goal of spreading deductive verification methods to a larger audience, and we are firmly engaged into continuing such kind of actions in the future.

5. Highlights of the Year

5.1. Highlights of the Year

- Martin Clochard has been awarded the GDR GPL 2018 prize (http://gdr-gpl.cnrs.fr/node/361) for his thesis entitled “Methods and tools for specification and proof of difficult properties of sequential programs” carried out at LRI, under the scientific supervision of Claude Marché and Andrei Paskevich. [57] [43]

  Martin Clochard is currently a postdoc at ETH Zurich.
- Jean-Christophe Filliâtre receives the 2019 CAV Award, jointly with Rustan Leino (Amazon Web Services), for the design and development of reusable intermediate verification languages which significantly simplified and accelerated the building of automated deductive verifiers. Jean-Christophe is the initial designer of the Why environment for automated deductive verification, and a leading developer of its successor Why3.

  The CAV award is given anually at the CAV conference for fundamental contributions to the field of Computer-Aided Verification. http://cavconference.org/cav-award
- Claude Marché received the FIEEC CARNOT 2019 prize for applied research for his collaboration with AdaCore. The award recognizes his collaboration with AdaCore, a computer assisted proof verification company, for applications in the development of critical software for safety and security in the aeronautics, space, air traffic control and rail transportation industries, autonomous vehicles, finance or medical devices. https://www.instituts-carnot.eu/fr/actualite/prix-fieec-carnot-de-la-recherche-appliquee-trois-chercheurs-recompenses-pour-leurs-partenariats-retd-avec-les-pme

6. New Software and Platforms

6.1. Alt-Ergo

Automated theorem prover for software verification

KEYWORDS: Software Verification - Automated theorem proving

FUNCTIONAL DESCRIPTION: Alt-Ergo is an automatic solver of formulas based on SMT technology. It is especially designed to prove mathematical formulas generated by program verification tools, such as Frama-C for C programs, or SPARK for Ada code. Initially developed in Toccata research team, Alt-Ergo's distribution and support are provided by OCamlPro since September 2013.

RELEASE FUNCTIONAL DESCRIPTION: the "SAT solving" part can now be delegated to an external plugin, new experimental SAT solver based on mini-SAT, provided as a plugin. This solver is, in general, more efficient on ground problems, heuristics simplification in the default SAT solver and in the matching (instantiation) module, re-implementation of internal literals representation, improvement of theories combination architecture, rewriting some parts of the formulas module, bugfixes in records and numbers modules, new option "-no-Ematching" to perform matching without equality reasoning (i.e. without considering "equivalence classes"). This option is very useful for benchmarks coming from Atelier-B, two new experimental options: "-save-used-context" and "-replay-used-context". When the goal is proved valid, the first option allows to save the names of useful axioms into a ".used" file. The second one is used to replay the proof using only the axioms listed in the corresponding ".used" file. Note that the replay may fail because of the absence of necessary ground terms generated by useless axioms (that are not included in .used file) during the initial run.

- Participants: Alain Mebsout, Évelyne Contejean, Mohamed Iguernelala, Stéphane Lescuyer and Sylvain Conchon
- Partner: OCamlPro
- Contact: Sylvain Conchon
- URL: http://alt-ergo.lri.fr

6.2. CoqInterval

Interval package for Coq

KEYWORDS: Interval arithmetic - Coq

FUNCTIONAL DESCRIPTION: CoqInterval is a library for the proof assistant Coq.

It provides several tactics for proving theorems on enclosures of real-valued expressions. The proofs are performed by an interval kernel which relies on a computable formalization of floating-point arithmetic in Coq.

The Marelle team developed a formalization of rigorous polynomial approximation using Taylor models in Coq. In 2014, this library has been included in CoqInterval.

- Participants: Assia Mahboubi, Érik Martin-Dorel, Guillaume Melquiond, Jean-Michel Muller, Laurence Rideau, Laurent Théry, Micaela Mayero, Mioara Joldes, Nicolas Brisebarre and Thomas Sibut-Pinote
- Contact: Guillaume Melquiond
- Publications: Proving bounds on real-valued functions with computations - Floating-point arithmetic in the Coq system - Proving Tight Bounds on Univariate Expressions with Elementary Functions in Coq - Formally Verified Approximations of Definite Integrals - Formally Verified Approximations of Definite Integrals
- URL: http://coq-interval.gforge.inria.fr/

6.3. Coquelicot

The Coquelicot library for real analysis in Coq

KEYWORDS: Coq - Real analysis

FUNCTIONAL DESCRIPTION: Coquelicot is library aimed for supporting real analysis in the Coq proof assistant. It is designed with three principles in mind. The first is the user-friendliness, achieved by implementing methods of automation, but also by avoiding dependent types in order to ease the stating and readability of theorems. This latter part was achieved by defining total function for basic operators, such as limits or integrals. The second principle is the comprehensiveness of the library. By experimenting on several applications, we ensured that the available theorems are enough to cover most cases. We also wanted to be able to extend our library towards more generic settings, such as complex analysis or Euclidean spaces. The third principle is for the Coquelicot library to be a conservative extension of the Coq standard library, so that it can be easily combined with existing developments based on the standard library.

- Participants: Catherine Lelay, Guillaume Melquiond and Sylvie Boldo
- Contact: Sylvie Boldo
- URL: http://coquelicot.saclay.inria.fr/

6.4. Cubicle

The Cubicle model checker modulo theories

KEYWORDS: Model Checking - Software Verification

FUNCTIONAL DESCRIPTION: Cubicle is an open source model checker for verifying safety properties of array-based systems, which corresponds to a syntactically restricted class of parametrized transition systems with states represented as arrays indexed by an arbitrary number of processes. Cache coherence protocols and mutual exclusion algorithms are typical examples of such systems.

- Participants: Alain Mebsout and Sylvain Conchon
- Contact: Sylvain Conchon
- URL: http://cubicle.lri.fr/

6.5. Flocq

The Flocq library for formalizing floating-point arithmetic in Coq

KEYWORDS: Floating-point - Arithmetic code - Coq

FUNCTIONAL DESCRIPTION: The Flocq library for the Coq proof assistant is a comprehensive formalization of floating-point arithmetic: core definitions, axiomatic and computational rounding operations, high-level properties. It provides a framework for developers to formally verify numerical applications.

Flocq is currently used by the CompCert verified compiler to support floating-point computations.

- Participants: Guillaume Melquiond, Pierre Roux and Sylvie Boldo
- Contact: Sylvie Boldo
- Publications: Flocq: A Unified Library for Proving Floating-point Algorithms in Coq - A Formally-Verified C Compiler Supporting Floating-Point Arithmetic - Verified Compilation of Floating-Point Computations - Innocuous Double Rounding of Basic Arithmetic Operations - Formal Proofs of Rounding Error Bounds : With application to an automatic positive definiteness check - Computer Arithmetic and Formal Proofs : Verifying Floating-point Algorithms with the Coq System
- URL: http://flocq.gforge.inria.fr/

6.6. Gappa

The Gappa tool for automated proofs of arithmetic properties

KEYWORDS: Floating-point - Arithmetic code - Software Verification - Constraint solving

FUNCTIONAL DESCRIPTION: Gappa is a tool intended to help formally verifying numerical programs dealing with floating-point or fixed-point arithmetic. It has been used to write robust floating-point filters for CGAL and it is used to verify elementary functions in CRlibm. While Gappa is intended to be used directly, it can also act as a backend prover for the Why3 software verification plateform or as an automatic tactic for the Coq proof assistant.

- Participant: Guillaume Melquiond
- Contact: Guillaume Melquiond

- Publications: Generating formally certified bounds on values and round-off errors - Formal certification of arithmetic filters for geometric predicates - Assisted verification of elementary functions - From interval arithmetic to program verification - Formally Certified Floating-Point Filters For Homogeneous Geometric Predicates - Combining Coq and Gappa for Certifying Floating-Point Programs - Handbook of Floating-Point Arithmetic - Certifying the floating-point implementation of an elementary function using Gappa - Automations for verifying floating-point algorithms - Automating the verification of floating-point algorithms - Computer Arithmetic and Formal Proofs : Verifying Floating-point Algorithms with the Coq System
- URL: http://gappa.gforge.inria.fr/

6.7. Why3

The Why3 environment for deductive verification

KEYWORDS: Formal methods - Trusted software - Software Verification - Deductive program verification

FUNCTIONAL DESCRIPTION: Why3 is an environment for deductive program verification. It provides a rich language for specification and programming, called WhyML, and relies on external theorem provers, both automated and interactive, to discharge verification conditions. Why3 comes with a standard library of logical theories (integer and real arithmetic, Boolean operations, sets and maps, etc.) and basic programming data structures (arrays, queues, hash tables, etc.). A user can write WhyML programs directly and get correct-by-construction OCaml programs through an automated extraction mechanism. WhyML is also used as an intermediate language for the verification of C, Java, or Ada programs.

- Participants: Andriy Paskevych, Claude Marché, François Bobot, Guillaume Melquiond, Jean-Christophe Filliâtre, Levs Gondelmans and Martin Clochard
- Partners: CNRS - Université Paris-Sud
- Contact: Claude Marché
- URL: http://why3.lri.fr/

6.8. Coq

The Coq Proof Assistant

KEYWORDS: Proof - Certification - Formalisation

SCIENTIFIC DESCRIPTION: Coq is an interactive proof assistant based on the Calculus of (Co-)Inductive Constructions, extended with universe polymorphism. This type theory features inductive and co-inductive families, an impredicative sort and a hierarchy of predicative universes, making it a very expressive logic. The calculus allows to formalize both general mathematics and computer programs, ranging from theories of finite structures to abstract algebra and categories to programming language metatheory and compiler verification. Coq is organised as a (relatively small) kernel including efficient conversion tests on which are built a set of higher-level layers: a powerful proof engine and unification algorithm, various tactics/decision procedures, a transactional document model and, at the very top an IDE.

FUNCTIONAL DESCRIPTION: Coq provides both a dependently-typed functional programming language and a logical formalism, which, altogether, support the formalisation of mathematical theories and the specification and certification of properties of programs. Coq also provides a large and extensible set of automatic or semi-automatic proof methods. Coq's programs are extractible to OCaml, Haskell, Scheme, ...

RELEASE FUNCTIONAL DESCRIPTION: Coq version 8.10 contains two major new features: support for a native fixed-precision integer type and a new sort SProp of strict propositions. It is also the result of refinements and stabilization of previous features, deprecations or removals of deprecated features, cleanups of the internals of the system and API, and many documentation improvements. This release includes many user-visible changes, including deprecations that are documented in the next subsection, and new features that are documented in the reference manual.

Version 8.10 is the fifth release of Coq developed on a time-based development cycle. Its development spanned 6 months from the release of Coq 8.9. Vincent Laporte is the release manager and maintainer of this release. This release is the result of 2500 commits and 650 PRs merged, closing 150+ issues.

See the Zenodo citation for more information on this release: https://zenodo.org/record/3476303#.Xe54f5NKjOQ

NEWS OF THE YEAR: Coq 8.10.0 contains:

- some quality-of-life bug fixes, - a critical bug fix related to template polymorphism, - native 63-bit machine integers, - a new sort of definitionally proof-irrelevant propositions: SProp, - private universes for opaque polymorphic constants, - string notations and numeral notations, - a new simplex-based proof engine for the tactics lia, nia, lra and nra, - new introduction patterns for SSReflect, - a tactic to rewrite under binders: under, - easy input of non-ASCII symbols in CoqIDE, which now uses GTK3.

All details can be found in the user manual.

- Participants: Yves Bertot, Frédéric Besson, Maxime Denes, Emilio Jesús Gallego Arias, Gaëtan Gilbert, Jason Gross, Hugo Herbelin, Assia Mahboubi, Érik Martin-Dorel, Guillaume Melquiond, Pierre-Marie Pédrot, Michael Soegtrop, Matthieu Sozeau, Enrico Tassi, Laurent Théry, Théo Zimmermann, Theo Winterhalter, Vincent Laporte, Arthur Charguéraud, Cyril Cohen, Christian Doczkal and Chantal Keller
- Partners: CNRS - Université Paris-Sud - ENS Lyon - Université Paris-Diderot
- Contact: Matthieu Sozeau
- URL: http://coq.inria.fr/

7. New Results

7.1. Foundations and Spreading of Deductive Program Verification

A Why3 Framework for Reflection Proofs and its Application to GMP's Algorithms Earlier works using Why3 showed that automatically verifying the algorithms of the arbitrary-precision integer library GMP exceeds the current capabilities of automatic solvers. To complete this verification, numerous cut indications had to be supplied by the user, slowing the project to a crawl. G. Melquiond and R. Rieu-Helft extended Why3 with a framework for proofs by reflection, with minimal impact on the trusted computing base. This framework makes it easy to write dedicated decision procedures that make full use of Why3's imperative features and are formally verified. This approach opens the way to efficiently tackling the further verification of GMP's algorithms [33].

GOSPEL - Providing OCaml with a Formal Specification Language In the context of the ANR project "VOCaL" (see Sec. 9.2.2), which aims at building a formally verified OCaml library of data structures and algorithms, a specification language for OCaml is designed and implemented. It is called GOSPEL, for Generic Ocaml SPEcification Language. During his post-doc in the Toccata team, from September 2018 to August 2019, C. Lourenço implemented a parser and type checker for GOSPEL. The work on the GOSPEL language has been presented at FM'19 [23]. J.-C. Filliâtre was keynote speaker at iFM 2019 and he gave a talk on the on-going work in the VOCaL project, including GOSPEL [17].

Program Verification Competition VerifyThis 2018 VerifyThis 2018 took place on April 14 and 15, 2018 in Thessaloniki, Greece, as a satellite event of ETAPS 2018. It was the sixth edition in the VerifyThis annual competition series. Typical challenges in the VerifyThis competitions are small but intricate algorithms given in pseudo-code with an informal specification in natural language. Participants have to formalize the requirements, implement a solution, and formally verify the implementation for adherence to the specification. There are no restrictions on the programming language and verification technology used. The time frame to solve each challenge is limited to 90 minutes. Submissions are judged for correctness, completeness, and elegance. The focus includes the usability of the tools, their facilities for formalizing the properties and providing helpful output.

VerifyThis 2018 consisted of three increasingly difficult verification challenges, selected to showcase various aspects of software verification. Eleven teams (one or two participants) took part in the competition. A full report on the VerifyThis 2018 event [39] provides a presentation of the competing teams, a description of the challenges, a high-level overview of the solutions, and the results of the competition.

Proof automation with the Coq proof assistant Proof assistants based on Type Theory, such as Coq, allow the implementation of effective automatic tactics based on computational reflection. These are usually limited to a particular mathematical domain (such as linear arithmetic or ring theory). In contrast, SMTCoq is a modular and extensible tool, using external provers, which generalizes these computational approaches to combine the reasoning from multiple domains. For this, it is based on a high-level interface, which offers greater expressiveness, at the cost of more complex automation. Q. Garchery and his co-authors [22] focused on two improvements: the ability to use quantified lemmas, and the ability to use multiple representations of the same data structure. They realized a new automatic tactic, based on SMTCoq, that is expressive while keeping the modularity and the efficiency of the latter. Such a tactic thus enable scalable, low-cost automation of new domains supported by state-of-the-art automatic provers.

Certificates for Logic Transformations In a context of formal program verification, using automatic provers, the trusted code base of verification environments is typically very broad. An environment such as Why3 implements many complex procedures: generation of verification conditions, logical transformations of proof tasks, and interactions with external provers. Considering only the logical transformations of Why3, their implementation already amounts to more than 17,000 lines of OCaml code. In order to increase our confidence in the correction of such a verification tool, Garchery, Keller, Marché and Paskevich present [32] proposed a mechanism of certifying transformations, producing certificates that can be validated by an external tool, according to the *skeptical* approach. They explored two methods to validate certificates: one based on a dedicated verifier developed in OCaml, the other based on the universal proof checker Dedukti. A specificity of their certificates is to be "small grains" and composable, which makes the approach incremental, allowing to gradually add new certifying transformations.

Reasoning About Universal Cubes in MCMT The Model Checking Modulo Theories (MCMT) framework is a powerful model checking technique for verifying safety properties of parameterized transition systems. In MCMT, logical formulas are used to represent both transitions and sets of states and safety properties are verified by an SMT-based backward reachability analysis. To be fully automated, the class of formulas handled in MCMT is restricted to cubes, i.e. existentially quantified conjunction of literals. While being very expressive, cubes cannot define properties with a global termination condition, usually described by a universally quantified formula. In this work, S. Conchon and M. Roux describe BRWP, an extension of the backward reachability of MCMT for reasoning about validity properties expressed as universal cubes, that is formulas of the form $\exists i \forall j. C(i, j)$, where $C(i, j)$ is a conjunction of literals. Their approach consists in a tight cooperation between the backward reachability loop and a deductive verification engine based on weakest-precondition calculus (WP). To provide evidence for the applicability of this new algorithm, they show how to make the model checker Cubicle cooperate with Why3 [25].

A Generalized Program Verification Workflow Based on Loop Elimination and SA Form.
C. Lourenço, together with Maria Frade and Jorge Sousa Pinto from Universidade do Minho, developed a minimal model of the functioning of program verification and property checking tools based on (i) the encoding of loops as non-iterating programs, either conservatively, making use of invariants and assume/assert commands, or in a bounded way; and (ii) the use of an intermediate single-assignment (SA) form. The model captures the basic workflow of tools like Boogie, Why3, or CBMC, building on a clear distinction between operational and axiomatic semantics. This allows one to consider separately the soundness of program annotation, loop encoding, translation into SA

form, and verification condition (VC) generation, as well as appropriate notions of completeness for each of these processes. To the best of our knowledge, this is the first formalization of a bounded model checking of software technique, including soundness and completeness proofs using Hoare logic; they also give the first completeness proof of a deductive verification technique based on a conservative encoding of invariant-annotated loops with assume/assert in SA form, as well as the first soundness proof based on a program logic. [21]

7.2. Reasoning on mutable memory in program verification

Certified Symbolic Execution Engine using Ghost Code Symbolic execution amounts to representing sets of concrete program states as a logical formula relating the program variables, and interpreting sets of executions as a transformation of that formula. B. Becker and C. Marché formalised the correctness of a symbolic interpreter engine, expressed by an over-approximation property stating that symbolic execution covers all concrete executions, and an under-approximation property stating that no useless symbolic states are generated. This formalisation is tailored for automated verification, that is the automated discharge of verification conditions to SMT solvers. To achieve this level of automation, they appropriately annotated the code of the symbolic interpreter with an original use of both ghost data and ghost statements [20].

Ghost Monitors M. Clochard, C. Marché and A. Paskevich proposed a new approach to deductive program verification based on auxiliary programs called *ghost monitors*. This technique is useful when the syntactic structure of the target program is not well suited for verification, for example, when an essentially recursive algorithm is implemented in an iterative fashion. This new approach consists in implementing, specifying, and verifying an auxiliary program that monitors the execution of the target program, in such a way that the correctness of the monitor entails the correctness of the target. The ghost monitor maintains the necessary data and invariants to facilitate the proof. It can be implemented and verified in any suitable framework, which does not have to be related to the language of the target programs. This technique is also applicable when one wants to establish relational properties between two target programs written in different languages and having different syntactic structure.

Ghost monitors can be used to specify and prove fine-grained properties about the *infinite behaviors* of target programs. Since this cannot be easily done using existing verification frameworks, this work introduces a dedicated language for ghost monitors, with an original construction to *catch* and handle divergent executions. The soundness of the underlying program logic is established using a particular flavor of transfinite games. This language and its soundness are formalized and mechanically checked. [24]

7.3. Verification of Computer Arithmetic

Formal Verification of a State-of-the-Art Integer Square Root Even though some programs only use integer operations, the best way to understand and verify them might be to view them as fixed-point arithmetic algorithm. This is the case of the function from the GMP library that computes the square root of a 64-bit integer. The C code is short but intricate, as it implements Newton's method and it relies on magic constants and intentional arithmetic overflows. G. Melquiond and R. Rieu-Helft have verified this algorithm using the Why3 tool and automated solvers such as Gappa [28].

Round-off error and exceptional behavior analysis of explicit Runge-Kutta methods S. Boldo, F. Faissole, and A. Chapoutot developed a new fine-grained analysis of round-off errors in explicit Runge-Kutta integration methods, taking into account exceptional behaviors, such as underflow and overflow [12]. First steps towards the formalization has been done by F. Faissole [34].

Optimal Inverse Projection of Floating-Point Addition In a setting where we have intervals for the values of floating-point variables x, a, and b, we are interested in improving these intervals when the floating-point equality $x \oplus a = b$ holds. This problem is common in constraint propagation, and called the inverse projection of the addition. D. Gallois-Wong, S. Boldo, and P. Cuoq proposed floating-point theorems that provide optimal bounds for all the intervals [13].

Emulating round-to-nearest-ties-to-zero "augmented" floating- point operations using round-to-nearest-ties-to-even ari The 2019 version of the IEEE 754 Standard for Floating-Point Arithmetic recommends that new "augmented" operations should be provided for the binary formats. These operations use a new "rounding direction": round to nearest *ties-to-zero*. S. Boldo, C. Lauter, and J.-M. Muller show how they can be implemented using the currently available operations, using round-to-nearest *ties-to-even* with a partial formal proof of correctness [42].

LTI filters Several developments were made towards the efficency and accuracy of the implementation of LTI (linear time-invariant) numerical filters: a word-length optimization problem under accuracy constraints [26] by T. Hilaire, H. Ouzia, and B. Lopez, and a tight worst-case error analysis [16] by A. Volkova, T. Hilaire, and C. Lauter.

7.4. Spreading Formal Proofs

7.4.1. Real Analysis

Formally Verified Approximations of Definite Integrals The CoqInterval library provides some tactics for computing and formally verifying numerical approximations of real-valued expressions inside the Coq system. In particular, it is able to compute reliable bounds on proper definite integrals [64]. A. Mahboubi, G. Melquiond, and T. Sibut-Pinote extended these algorithms to also cover some improper integrals, e.g., those with an unbounded integration domain [14]. This makes CoqInterval one of the very few tools able to produce reliable results for improper integrals, be they formally verified or not.

Coq Formalization of algorithms for numerical filters D. Gallois-Wong developed a Coq formalization of a generic representation of numerical filters, called SIF [31] in order to encompass all other representations of filters, and prove useful theorems only once.

Complexity theory and constructive analysis E. Neumann and F. Steinberg extended the framework for complexity of operators in analysis devised by Kawamura and Cook (2012) to allow for the treatment of a wider class of representations and applied it to the study of interval computation [15]. A. Kawamura, F. Steinberg, and H. Thies put forward a complexity class of type-two linear-time [27].

F. Steinberg, L. Théry, and H. Thies give a number of formal proofs of theorems from the field of computable analysis. Results include that the algebraic operations and the efficient limit operator on the reals are computable, that certain countably infinite products are isomorphic to spaces of functions, compatibility of the enumeration representation of subsets of natural numbers with the abstract definition of the space of open subsets of the natural numbers, and that continuous realizability implies sequential continuity [46] [29]. F. Steinberg and H. Thies formalized proofs about Baire spaces and the isomorphy of the concrete and abstract spaces of open sets [45].

7.4.2. Formal Analysis of Debian packages

Several new results were produced in the context of the CoLiS project for the formal analysis of Debian packages. A first important step is the version 2 of the design of the CoLiS language done by B. Becker, C. Marché and other co-authors [38], that includes a modified formal syntax, a extended formal semantics, together with the design of concrete and symbolic interpreters. Those interpreters are specified and implemented in Why3, proved correct (following the initial approach for the concrete interpreter published in 2018 [60] and the recent approach for symbolic interpretation mentioned above [20]), and finally extracted to OCaml code.

To make the extracted code effective, it must be linked together with a library that implements a solver for feature constraints [61], and also a library that formally specifies the behavior of basic UNIX utilities. The latter library is documented in details in a research report [40].

A third result is a large verification campaign running the CoLiS toolbox on all the packages of the current Debian distribution. The results of this campaign were reported in another article [41] that will be presented at TACAS conference in 2020. The most visible side effect of this experiment is the discovery of bugs: more than 150 bugs report have been filled against various Debian packages.

7.4.3. Miscellaneous

Functional Programming. J.-C. Filliâtre was invited speaker at JFLA 2019, as part of a session celebrating the 30 years of JFLA (a French-speaking national conference related to functional programming). He talked about 25 years of programming with OCaml [18]. At JFLA 2020, J.-C. Filliâtre will give a talk related to the elimination of non-tail calls [30].

Formal Verification of "ParcourSup" algorithms. In May–July 2019, Léo Andrès (M1 student at Paris Sud) did a three month internship on the verification of the first algorithm of Parcoursup using Why3. Most of the expected properties, taken from the public description of Parcoursup's algorithms, have been verified. Léo Andrès's report (in French), is available on-line [37]. In June-December 2019, Benedikt Becker worked on the verification of the Java source code of ParcourSup. The findings and lessons learnt are described in a report under preparation.

Formalizing loop-carried dependencies in Coq for high-level synthesis. F. Faissole, G. Constantinides, and D. Thomas developed Coq formalizations in order to improve high-level synthesis for FPGAs [44].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

We have two bilateral contracts which are closely related to a joint effort called the ProofInUse joint Laboratory. The objective of ProofInUse is to provide verification tools, based on mathematical proof, to industry users. These tools are aimed at replacing or complementing the existing test activities, whilst reducing costs.

This joint laboratory is a follow-up of the former "LabCom ProofInUse" between Toccata and the SME AdaCore, funded by the ANR programme "Laboratoires communs", from April 2014 to March 2017 http://www.spark-2014.org/proofinuse.

8.1.1. ProofInUse-AdaCore Collaboration

Participants: Claude Marché [contact], Jean-Christophe Filliâtre, Andrei Paskevich, Guillaume Melquiond, Sylvain Dailler.

This collaboration is a joint effort of the Inria project-team Toccata and the AdaCore company which provides development tools for the Ada programming language. It is funded by a 5-year bilateral contract from Jan 2019 to Dec 2023.

The SME AdaCore is a software publisher specializing in providing software development tools for critical systems. A previous successful collaboration between Toccata and AdaCore enabled Why3 technology to be put into the heart of the AdaCore-developed SPARK technology.

The objective of ProofInUse-AdaCore is to significantly increase the capabilities and performances of the Spark/Ada verification environment proposed by AdaCore. It aims at integration of verification techniques at the state-of-the-art of academic research, via the generic environment Why3 for deductive program verification developed by Toccata.

8.1.2. ProofInUse-MERCE Collaboration

Participants: Claude Marché [contact], Jean-Christophe Filliâtre, Andrei Paskevich, Guillaume Melquiond, Sylvain Dailler.

This bilateral contract is part of the ProofInUse effort. This collaboration joins efforts of the Inria project-team Toccata and the company Mitsubishi Electric R&D (MERCE) in Rennes. It is funded by a bilateral contract of 18 months from Nov 2019 to April 2021.

MERCE has strong and recognized skills in the field of formal methods. In the industrial context of the Mitsubishi Electric Group, MERCE has acquired knowledge of the specific needs of the development processes and meets the needs of the group in different areas of application by providing automatic verification and demonstration tools adapted to the problems encountered.

The objective of ProofInUse-MERCE is to significantly improve on-going MERCE tools regarding the verification of Programmable Logic Controllers and also regarding the verification of numerical C codes.

8.2. Bilateral Grants with Industry

8.2.1. CIFRE contract with TrustInSoft company

Participants: Guillaume Melquiond [contact], Raphaël Rieu-Helft.

Jointly with the thesis of R. Rieu-Helft, supervised in collaboration with the TrustInSoft company, we established a 3-year bilateral collaboration contract, that started in October 2017. The aim is to design methods that make it possible to design an arbitrary-precision integer library that, while competitive with the state-of-the-art library GMP, is formally verified. Not only are GMP's algorithm especially intricate from an arithmetic point of view, but numerous tricks were also used to optimize them. We are using the Why3 programming language to implement the algorithms, we are developing reflection-based procedures to verify them, and we finally extract them as a C library that is binary-compatible with GMP [9] [67] [33].

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. ELEFFAN

Participant: Sylvie Boldo [contact].

ELEFFAN is a Digicosme project funding the PhD of F. Faissole. S. Boldo is the principal investigator. It began in 2016 for three years. https://project.inria.fr/eleffan/

The ELEFFAN project aims at formally proving rounding error bounds of numerical schemes.

Partners: ENSTA Paristech (A. Chapoutot)

9.1.2. MILC

Participant: Sylvie Boldo [contact].

MILC is a DIM-RFSI project. It is a one-year project (2018–2019) that aims at formalizing measure theory and Lebesgue integral in the Coq proof assistant. https://lipn.univ-paris13.fr/MILC/

Partners: Université Paris 13 (M. Mayero, PI), Inria Paris, Inria Saclay

9.2. National Initiatives

9.2.1. ANR CoLiS

Participants: Claude Marché [contact], Andrei Paskevich.

The CoLiS research project is funded by the programme "Société de l'information et de la communication" of the ANR, for a period of 60 months, starting on October 1st, 2015. http://colis.irif.univ-paris-diderot.fr/

The project aims at developing formal analysis and verification techniques and tools for scripts. These scripts are written in the POSIX or bash shell language. Our objective is to produce, at the end of the project, formal methods and tools allowing to analyze, test, and validate scripts. For this, the project will develop techniques and tools based on deductive verification and tree transducers stemming from the domain of XML documents.

Partners: Université Paris-Diderot, IRIF laboratory (formerly PPS & LIAFA), coordinator; Inria Lille, team LINKS

9.2.2. ANR Vocal

Participants: Jean-Christophe Filliâtre [contact], Andrei Paskevich.

The Vocal research project is funded by the programme "Société de l'information et de la communication" of the ANR, for a period of 60 months, starting on October 1st, 2015. See https://vocal.lri.fr/

The goal of the Vocal project is to develop the first formally verified library of efficient general-purpose data structures and algorithms. It targets the OCaml programming language, which allows for fairly efficient code and offers a simple programming model that eases reasoning about programs. The library will be readily available to implementers of safety-critical OCaml programs, such as Coq, Astrée, or Frama-C. It will provide the essential building blocks needed to significantly decrease the cost of developing safe software. The project intends to combine the strengths of three verification tools, namely Coq, Why3, and CFML. It will use Coq to obtain a common mathematical foundation for program specifications, as well as to verify purely functional components. It will use Why3 to verify a broad range of imperative programs with a high degree of proof automation. Finally, it will use CFML for formal reasoning about effectful higher-order functions and data structures making use of pointers and sharing.

Partners: team Gallium (Inria Paris-Rocquencourt), team DCS (Verimag), TrustInSoft, and OCamlPro.

9.2.3. FUI LCHIP

Participant: Sylvain Conchon [contact].

LCHIP (Low Cost High Integrity Platform) is aimed at easing the development of safety critical applications (up to SIL4) by providing: (i) a complete IDE able to automatically generate and prove bounded complexity software (ii) a low cost, safe execution platform. The full support of DSLs and third party code generators will enable a seamless deployment into existing development cycles. LCHIP gathers scientific results obtained during the last 20 years in formal methods, proof, refinement, code generation, etc. as well as a unique return of experience on safety critical systems design. http://www.clearsy.com/en/2016/10/4260/

Partners: 2 technology providers (ClearSy, OcamlPro), in charge of building the architecture of the platform; 3 labs (IFSTTAR, LIP6, LRI), to improve LCHIP IDE features; 2 large companies (SNCF, RATP), representing public ordering parties, to check compliance with standard and industrial railway use-case.

The project lead by ClearSy has started in April 2016 and lasts 3 years. It is funded by BpiFrance as well as French regions.

9.2.4. ANR PARDI

Participant: Sylvain Conchon [contact].

Verification of PARameterized DIstributed systems. A parameterized system specification is a specification for a whole class of systems, parameterized by the number of entities and the properties of the interaction, such as the communication model (synchronous/asynchronous, order of delivery of message, application ordering) or the fault model (crash failure, message loss). To assist and automate verification without parameter instantiation, PARDI uses two complementary approaches. First, a fully automatic model checker modulo theories is considered. Then, to go beyond the intrinsic limits of parameterized model checking, the project advocates a collaborative approach between proof assistant and model checker. http://pardi.enseeiht.fr/

The proof lead by Toulouse INP/IRIT started in 2016 and lasts for 4 years. Partners: Université Pierre et Marie Curie (LIP6), Université Paris-Sud (LRI), Inria Nancy (team VERIDIS)

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. EMC2

Participant: Sylvie Boldo [contact].

A new ERC Synergy Grant 2018 project, called Extreme-scale Mathematically-based Computational Chemistry (EMC2) has just been accepted. The PIs are É. Cancès, L. Grigori, Y. Maday and J.-P. Piquemal. S. Boldo is part of the work package 3: validation and certification of molecular simulation results. https://www.sorbonne-universite.fr/newsroom/actualites/erc-synergy-grant-2018

9.3.2. Collaborations in European Programs, Except FP7 & H2020

Program: COST (European Cooperation in Science and Technology).

Project acronym: EUTypes https://eutypes.cs.ru.nl/

Project title: The European research network on types for programming and verification

Duration: 2015-2019

Coordinator: Herman Geuvers, Radboud University Nijmegen, The Netherlands

Other partners: 36 members countries, see http://www.cost.eu/COST_Actions/ca/CA15123?parties

Abstract: Types are pervasive in programming and information technology. A type defines a formal interface between software components, allowing the automatic verification of their connections, and greatly enhancing the robustness and reliability of computations and communications. In rich dependent type theories, the full functional specification of a program can be expressed as a type. Type systems have rapidly evolved over the past years, becoming more sophisticated, capturing new aspects of the behaviour of programs and the dynamics of their execution.

This COST Action will give a strong impetus to research on type theory and its many applications in computer science, by promoting (1) the synergy between theoretical computer scientists, logicians and mathematicians to develop new foundations for type theory, for example as based on the recent development of "homotopy type theory", (2) the joint development of type theoretic tools as proof assistants and integrated programming environments, (3) the study of dependent types for programming and its deployment in software development, (4) the study of dependent types for verification and its deployment in software analysis and verification. The action will also tie together these different areas and promote cross-fertilisation.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Jorge Sousa Pinto, professor from Universidade do Minho (Braga, Portugal, https://haslab.uminho.pt/jsp/) visited the team for 1 month in May 2019. We interact with him on the topic of the formalization of VC generation algorithms [21]. He also proposed a formalization using the Why3 tool.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

J.-C. Filliâtre, scientific chair and co-organizer of EJCP (École Jeunes Chercheurs en Programmation du GDR GPL) at Strasbourg on June 24–28, 2019. 5 days / 8 lectures / 25 participants. http://ejcp2019.icube.unistra.fr/

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

S. Boldo, program co-chair of the 26th IEEE Symposium on Computer Arithmetic (ARITH 2019), Kyoto, Japan. [35]

10.1.2.2. Member of the Conference Program Committees

S. Boldo, 11th NASA Formal Methods Symposium (NFM 2019)

J.-C. Filliâtre, 10th International Conference on Interactive Theorem Proving (ITP 2019)

J.-C. Filliâtre, 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)

J.-C. Filliâtre, European Symposium on Programming (ESOP 2020)

G. Melquiond, 26th IEEE Symposium on Computer Arithmetic (ARITH 2019)

G. Melquiond, 10th International Conference on Interactive Theorem Proving (ITP 2019)

10.1.2.3. Reviewer

The members of the Toccata team have reviewed numerous papers for numerous international conferences.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

G. Melquiond, member of the editorial board of *Reliable Computing*.

J.-C. Filliâtre, member of the editorial board of *Journal of Functional Programming*.

10.1.3.2. Reviewer - Reviewing Activities

The members of the Toccata team have reviewed numerous papers for numerous international journals.

10.1.4. Invited Talks

G. Melquiond, "Computer Arithmetic and Formal Proofs" [19], at the 30th Journées Francophones des Langages Applicatifs (JFLA 2019).

J.-C. Filliâtre, "Deductive Verification of OCaml Libraries" [17], at the 15th International Conference on integrated Formal Methods (iFM), December 2019, Bergen, Norway.

J.-C. Filliâtre, "Retour sur 25 ans de programmation avec OCaml" [18], at Journées Francophones des Langages Applicatifs (JFLA), January 2019, Les Rousses, France.

J.-C. Filliâtre, "The Why3 tool for deductive verification and verified OCaml libraries", at Frama-C/Spark day 2019, Paris, France.

10.1.5. Leadership within the Scientific Community

S. Boldo, elected chair of the ARITH working group of the GDR-IM (a CNRS subgroup of computer science) with J. Detrey (Inria Nancy) and now L.-S. Didier (Univ. Toulon).

J.-C. Filliâtre, chair of IFIP WG 1.9/2.15 verified Software.

10.1.6. Scientific Expertise

G. Melquiond, member of the scientific commission of Inria-Saclay, in charge of selecting candidates for PhD grants, Post-doc grants, temporary leaves from universities ("délégations").

S. Boldo, member of the program committee for selecting postdocs of the maths/computer science program of the Labex mathématique Hadamard.

S. Boldo has mentored for 1 year a female PhD student of University Paris-Saclay (from epidemiology).

J.-C. Filliâtre, grading the entrance examination at X/ENS (*"option informatique"*).

C. Marché, member of DigiCosme committee for research and innovation (selection of projects for working groups, post-doc grants, doctoral missions, invited professors)

10.1.7. Research Administration

G. Melquiond, member of the committee for the monitoring of PhD students (*"commission de suivi doctoral"*).

S Boldo, member of the CLFP (*"commission locale de formation permanente"*).

S. Boldo, member of the (national) IES commission.

S. Boldo, member of the CDT commission of Saclay (*"commission de développement technologique"*).

S. Boldo, member of the STIC department commission of Univ. Paris-Sud.

S. Boldo, member of the executive commission for the Digicosme Labex.

S. Boldo, deputy scientific director (DSA) of Inria Saclay research center from January 1st, 2019

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master Parisien de Recherche en Informatique (MPRI) https://wikimpri.dptinfo.ens-cachan.fr/doku.php: "Proofs of Programs" http://www.lri.fr/~marche/MPRI-2-36-1/ (M2), C. Marché (12h).

A. Lanco, *Unix et programmation Web*, 24h, L2, Université Paris-Saclay, France.

G. Melquiond, *Programmation C++ avancée*, 12h, M2, Université Paris-Saclay, France.

J.-C. Filliâtre, *Langages de programmation et compilation*, 25h, L3, École Normale Supérieure, France.

J.-C. Filliâtre, *Les bases de l'algorithmique et de la programmation*, 15h, L3, École Polytechnique, France.

J.-C. Filliâtre, *Compilation*, 18h, M1, École Polytechnique, France.

S. Boldo was invited in the Xavier Leroy's course at Collège de France on Dec 19th.

10.2.2. Supervision

HDR: G. Melquiond, "Formal Verification for Numerical Computations, and the Other Way Around" [11], Université Paris-Sud, Apr. 1st, 2019.

PhD: F. Faissole, "Formalisations d'analyses d'erreurs en analyse numérique et en arithmétique à virgule flottante", Université Paris-Saclay & Université Paris-Sud, Dec 13th 2019, supervised by S. Boldo and A. Chapoutot. [Not yet on HAL]

PhD: A. Coquereau, "Amélioration de performances du solveur SMT Alt-Ergo grâce à l'intégration d'un solveur SAT efficace", Université Paris-Saclay & Université Paris-Sud, Dec 16th 2019, supervised by S. Conchon, F. Le Fessant et M. Mauny. [Not yet on HAL]

PhD: M. Roux, "Extensions de l'algorithme d'atteignabilité arrière dans le cadre de la vérification de modèles modulo théories", Université Paris-Saclay & Université Paris-Sud, Dec 19th 2019, supervised by Sylvain Conchon. [Not yet on HAL]

PhD in progress: R. Rieu-Helft, "Développement et vérification de bibliothèques d'arithmétique entière en précision arbitraire", since Oct. 2017, supervised by G. Melquiond and P. Cuoq (TrustInSoft).

PhD in progress: D. Gallois-Wong, “Vérification formelle et filtres numériques”, since Oct. 2017, supervised by S. Boldo and T. Hilaire.

PhD in progress: Q. Garchery, “Certification de la génération et de la transformation d’obligations de preuve”, since Oct. 2018, supervised by C. Keller, C. Marché and A. Paskevich.

PhD in progress: A. Lanco, “Stratégies pour la réduction forte”, since Oct. 2019, supervised by T. Balabonski and G. Melquiond.

G. Turbiau has been a M1 trainee for 2 months, supervised by S. Boldo.

10.2.3. Juries

S. Boldo, member of the Inria CRCN recruiting juries, national one and at Rennes.

S. Boldo, president of the PhD defense of Clothilde Jeangoudoux, Sorbonne University, May 21st

S. Boldo, reviewer of the habilitation of Christoph Lauter, Sorbonne University, May 22nd

S. Boldo, reviewer of the PhD defence of Florent Bréhard, ENS Lyon, July 12th

S. Boldo, reviewer of PhD defence of Damien Rouhling, Université Côte d’Azur, Sept 30th

S. Boldo, president of the PhD defense of Yohan Chatelain, Université Paris-Saclay, Dec 12th

S. Boldo, president of the PhD defense of Armaël Guéneau, Université de Paris, Dec 16th

S. Boldo, member of the PhD defense of Antoine Kaszczyc, Université Paris-Nord, Dec 18th

J.-C. Filliâtre, reviewer of the PhD defense of Y. El Bakouny, “Scallina: On the Intersection of Scala and Gallina”, Université Saint-Joseph, Beirut, Lebanon, December 19, 2019.

J.-C. Filliâtre, president of the PhD defense of L. Blatter “Relational properties for specification and verification of C programs in Frama-C”, University Paris Saclay, September 26, 2019.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

S. Boldo has been the scientific head for Saclay for the MECSI group for networking about computer science popularization inside Inria until October 2019.

She was also responsible (with A. Couvreur of the Grace team a,d M. Quet of the SCM) for the 2019 “Fête de la science” on October 10th and 11th 2019. Teenagers were welcomed on 8 activities ranging from unplugged activities with Duplo construction toys to programming, and from applied mathematics to theoretical computer science.

10.3.2. Education

S. Conchon and J.-C. Filliâtre, together with K. Nguyen and T. Balabonski (members of the LRI team but not members of Toccata), wrote a book “Numérique et Sciences Informatiques, 30 leçons avec exercices corrigés” (Ellipses, August 2019) [36] targeting the new Computer Science programme in the French high school system, which started in September 2019.

S. Boldo gave a talk and was involved in the organization of the “Rencontres des Jeunes mathématiciennes et Informaticiennes” organized in Saclay on Oct 21st and 22nd 2019. About 20 female teenagers were welcomed for two days with talks and exercises to discover research.

10.3.3. Interventions

C. Marché presented an overview of the Why3 tool and its ecosystem at the Paris Open Source Summit in Saint-Denis, Dec. 11, 2019, as part of a workshop “ATELIER Inria : OCaml, Coq, Why 3: pour concevoir codes et infrastructures plus sûrs en C, Java, Javascript, Ada” https://www.opensourcesummit.paris/conferences.html?date=1576022400).

S. Boldo gave a talk at a high school in Sceaux on Feb 8th.

S. Boldo presented the floating-point arithmetic at IRIF to the members of the CODYS ANR on April 1st.

11. Bibliography

Major publications by the team in recent years

[1] F. Bobot, J.-C. Filliâtre, C. Marché, A. Paskevich.*Let's Verify This with Why3*, in "International Journal on Software Tools for Technology Transfer (STTT)", 2015, vol. 17, n° 6, p. 709–727, http://hal.inria.fr/hal-00967132/en

[2] S. Boldo, J.-H. Jourdan, X. Leroy, G. Melquiond.*Verified Compilation of Floating-Point Computations*, in "Journal of Automated Reasoning", February 2015, vol. 54, n° 2, p. 135-163, https://hal.inria.fr/hal-00862689

[3] S. Boldo, G. Melquiond.*Computer Arithmetic and Formal Proofs: Verifying Floating-point Algorithms with the Coq System*, ISTE Press - Elsevier, December 2017, https://hal.inria.fr/hal-01632617

[4] M. Clochard, L. Gondelman, M. Pereira.*The Matrix Reproved*, in "Journal of Automated Reasoning", 2018, vol. 60, n° 3, p. 365–383, https://hal.inria.fr/hal-01617437

[5] S. Conchon, M. Iguernlala, K. Ji, G. Melquiond, C. Fumex.*A Three-tier Strategy for Reasoning about Floating-Point Numbers in SMT*, in "Computer Aided Verification", 2017, https://hal.inria.fr/hal-01522770

[6] J.-C. Filliâtre, L. Gondelman, A. Paskevich.*The Spirit of Ghost Code*, in "Formal Methods in System Design", 2016, vol. 48, n° 3, p. 152–174, https://hal.archives-ouvertes.fr/hal-01396864v1

[7] C. Fumex, C. Dross, J. Gerlach, C. Marché.*Specification and Proof of High-Level Functional Properties of Bit-Level Programs*, in "8th NASA Formal Methods Symposium", Minneapolis, MN, USA, S. Rayadurgam, O. Tkachuk (editors), Lecture Notes in Computer Science, Springer, June 2016, vol. 9690, p. 291–306, https://hal.inria.fr/hal-01314876

[8] É. Martin-Dorel, G. Melquiond.*Proving Tight Bounds on Univariate Expressions with Elementary Functions in Coq*, in "Journal of Automated Reasoning", 2016, https://hal.inria.fr/hal-01086460

[9] G. Melquiond, R. Rieu-Helft.*A Why3 Framework for Reflection Proofs and its Application to GMP's Algorithms*, in "9th International Joint Conference on Automated Reasoning", Oxford, United Kingdom, D. Galmiche, S. Schulz, R. Sebastiani (editors), Lecture Notes in Computer Science, July 2018, https://hal.inria.fr/hal-01699754

[10] P. Roux, M. Iguernlala, S. Conchon.*A Non-linear Arithmetic Procedure for Control-Command Software Verification*, in "24th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", Thessalonique, Greece, April 2018, https://hal.archives-ouvertes.fr/hal-01737737

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] G. MELQUIOND.*Formal Verification for Numerical Computations, and the Other Way Around*, Université Paris Sud, April 2019, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-02194683

Articles in International Peer-Reviewed Journal

[12] S. BOLDO, F. FAISSOLE, A. CHAPOUTOT.*Round-off error and exceptional behavior analysis of explicit Runge-Kutta methods*, in "IEEE Transactions on Computers", 2019, forthcoming [*DOI :* 10.1109/TC.2019.2917902], https://hal.archives-ouvertes.fr/hal-01883843

[13] D. GALLOIS-WONG, S. BOLDO, P. CUOQ.*Optimal Inverse Projection of Floating-Point Addition*, in "Numerical Algorithms", 2019, forthcoming [*DOI :* 10.1007/S11075-019-00711-Z], https://hal.inria.fr/hal-01939097

[14] A. MAHBOUBI, G. MELQUIOND, T. SIBUT-PINOTE.*Formally Verified Approximations of Definite Integrals*, in "Journal of Automated Reasoning", February 2019, vol. 62, n^{o} 2, p. 281-300 [*DOI :* 10.1007/S10817-018-9463-7], https://hal.inria.fr/hal-01630143

[15] E. NEUMANN, F. STEINBERG.*Parametrised second-order complexity theory with applications to the study of interval computation*, in "Theoretical Computer Science", 2019, forthcoming [*DOI :* 10.1016/J.TCS.2019.05.009], https://hal.inria.fr/hal-02148494

[16] A. VOLKOVA, T. HILAIRE, C. LAUTER.*Arithmetic approaches for rigorous design of reliable Fixed-Point LTI filters*, in "IEEE Transactions on Computers", 2019, p. 1-14, forthcoming [*DOI :* 10.1109/TC.2019.2950658], https://hal.archives-ouvertes.fr/hal-01918650

Invited Conferences

[17] J.-C. FILLIÂTRE.*Deductive Verification of OCaml Libraries*, in "iFM 2019 - International Conference on integrated Formal Methods", Bergen, Norway, December 2019, https://hal.inria.fr/hal-02406253

[18] J.-C. FILLIÂTRE.*Retour sur 25 ans de programmation avec OCaml*, in "JFLA 2019 - Journées Francophones des Langages Applicatifs", Les Rousses, France, January 2019, https://hal.inria.fr/hal-02406208

[19] G. MELQUIOND.*Computer Arithmetic and Formal Proofs*, in "30èmes Journées Francophones des Langages Applicatifs", Les Rousses, France, N. MAGAUD (editor), January 2019, https://hal.inria.fr/hal-02013540

International Conferences with Proceedings

[20] B. BECKER, C. MARCHÉ.*Ghost Code in Action: Automated Verification of a Symbolic Interpreter*, in "VSTTE 2019 - 11th Working Conference on Verified Software: Tools, Techniques and Experiments", New York, United States, S. CHAKRABORTY, J. A.NAVAS (editors), Lecture Notes in Computer Science, September 2019, https://hal.inria.fr/hal-02276257

[21] C. BELO LOURENÇO, M. J. FRADE, J. SOUSA PINTO.*A Generalized Program Verification Workflow Based on Loop Elimination and SA Form*, in "FormaliSE 2019 - 7th International Conference on Formal Methods in Software Engineering", Montreal, Canada, May 2019, https://hal.inria.fr/hal-02431769

[22] V. BLOT, A. BOUSALEM, Q. GARCHERY, C. KELLER.*SMTCoq: automatisation expressive et extensible dans Coq*, in "Journées Francophones des Langages Applicatifs", Les Rousses, France, January 2019, https://hal.archives-ouvertes.fr/hal-02369249

[23] A. CHARGUÉRAUD, J.-C. FILLIÂTRE, C. LOURENÇO, M. PEREIRA.*GOSPEL -Providing OCaml with a Formal Specification Language*, in "FM 2019 - 23rd International Symposium on Formal Methods", Porto, Portugal, October 2019, https://hal.inria.fr/hal-02157484

[24] M. CLOCHARD, C. MARCHÉ, A. PASKEVICH.*Deductive Verification with Ghost Monitors*, in "Principles of Programming Languages", New Orleans, United States, 2020 [*DOI :* 10.1145/3371070], https://hal.inria.fr/hal-02368284

[25] S. CONCHON, M. ROUX.*Reasoning about Universal Cubes in MCMT*, in "Formal Methods and Software Engineering", Shenzhen, China, Lecture Notes in Computer Science, 2019, n^{o} 11852, p. 270–285, https://hal.archives-ouvertes.fr/hal-02420588

[26] T. HILAIRE, H. OUZIA, B. LOPEZ.*Optimal Word-Length Allocation for the Fixed-Point Implementation of Linear Filters and Controllers*, in "2019 IEEE 26th Symposium on Computer Arithmetic (ARITH)", Kyoto, Japan, IEEE, June 2019, p. 175-182 [*DOI :* 10.1109/ARITH.2019.00040], https://hal.sorbonne-universite.fr/hal-02393851

[27] A. KAWAMURA, F. STEINBERG, H. THIES.*Second-Order Linear-Time Computability with Applications to Computable Analysis*, in "TAMC 2019 - 15th Annual Conference Theory and Applications of Models of Computation", Tokyo, Japan, T. V. GOPAL, J. WATADA (editors), LNCS - Lecture Notes in Computer Science, Springer, March 2019, vol. 11436, p. 337-358, https://hal.inria.fr/hal-02148490

[28] G. MELQUIOND, R. RIEU-HELFT.*Formal Verification of a State-of-the-Art Integer Square Root*, in "ARITH-26 2019 - 26th IEEE 26th Symposium on Computer Arithmetic", Kyoto, Japan, S. BOLDO, M. LANGHAMMER (editors), June 2019, p. 183-186 [*DOI :* 10.1109/ARITH.2019.00041], https://hal.inria.fr/hal-02092970

[29] F. STEINBERG, H. THIES, L. THÉRY.*Quantitative continuity and Computable Analysis in Coq*, in "ITP 2019 - Tenth International Conference on Interactive Theorem Proving", Portland, United States, 2019, The version accepted to the conference can be accessed at https://drops.dagstuhl.de/opus/volltexte/2019/11083/ [*DOI :* 10.4230/LIPICS.ITP.2019.28], https://hal.archives-ouvertes.fr/hal-02426470

National Conferences with Proceeding

[30] J.-C. FILLIÂTRE.*Mesurer la hauteur d'un arbre*, in "Journées Francophones des Langages Applicatifs", Gruissan, France, January 2020, https://hal.inria.fr/hal-02315541

[31] D. GALLOIS-WONG.*Formalisation en Coq d'algorithmes de filtres numériques*, in "Journées Francophones des Langages Applicatifs (JFLA) 2019", Les Rousses, France, Journées Francophones des Langages Applicatifs 2019, Nicolas Magaud, January 2019, https://hal.inria.fr/hal-01929531

[32] Q. GARCHERY, C. KELLER, C. MARCHÉ, A. PASKEVICH.*Des transformations logiques passent leur certicat*, in "Journées Francophones des Langages Applicatifs", Gruissan, France, 2020, https://hal.inria.fr/hal-02384946

[33] R. RIEU-HELFT.*Un mécanisme de preuve par réflexion pour Why3 et son application aux algorithmes de GMP*, in "JFLA 2019 - 30èmes Journées Francophones des Langages Applicatifs", Rousses, France, January 2019, https://hal.inria.fr/hal-01943010

Conferences without Proceedings

[34] F. FAISSOLE.*Formalisation en Coq des erreurs d'arrondi de méthodes de Runge-Kutta pour les systèmes matriciels*, in "AFADL 2019 - 18e journées Approches Formelles dans l'Assistance au Développement de Logiciels", Toulouse, France, June 2019, https://hal.archives-ouvertes.fr/hal-02391924

Scientific Books (or Scientific Book chapters)

[35] *Proceedings of the 26th Symposium on Computer Arithmetic (ARITH 2019)*, IEEE, Kyoto, June 2019 [*DOI :* 10.1109/ARITH.2019.00005], https://hal.inria.fr/hal-02433652

[36] T. BALABONSKI, S. CONCHON, J.-C. FILLIÂTRE, K. NGUYỄN.*Numérique et Sciences Informatiques, 30 leçons avec exercices corrigés*, Ellipses, August 2019, https://hal.inria.fr/hal-02379073

Research Reports

[37] L. ANDRÈS.*Vérification par preuve formelle de propriétés fonctionnelles d'algorithme de classification*, Université Paris Sud (Paris 11) - Université Paris Saclay, August 2019, https://hal.inria.fr/hal-02421484

[38] B. BECKER, N. JEANNEROD, C. MARCHÉ, R. TREINEN.*Revision 2 of CoLiS language: formal syntax, semantics, concrete and symbolic interpreters*, ANR, October 2019, https://hal.inria.fr/hal-02321743

[39] M. HUISMAN, R. MONAHAN, P. MÜLLER, A. PASKEVICH, G. ERNST.*VerifyThis 2018: A Program Verification Competition*, Université Paris-Saclay, January 2019, https://hal.inria.fr/hal-01981937

[40] N. JEANNEROD, C. MARCHÉ, Y. RÉGIS-GIANAS, M. SIGHIREANU, R. TREINEN.*Specification of UNIX Utilities*, ANR, October 2019, https://hal.inria.fr/hal-02321691

Other Publications

[41] B. BECKER, N. JEANNEROD, C. MARCHÉ, Y. RÉGIS-GIANAS, M. SIGHIREANU, R. TREINEN.*Analysing installation scenarios of Debian packages*, 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02355602

[42] S. BOLDO, C. Q. LAUTER, J.-M. MULLER.*Emulating round-to-nearest-ties-to-zero "augmented" floating-point operations using round-to-nearest-ties-to-even arithmetic*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02137968

[43] M. CLOCHARD.*Méthodes et outils pour la spécification et la preuve de propriétés difficiles de programmes séquentiels (Documents de soutenance)*, February 2019, https://hal.inria.fr/hal-02047458

[44] F. FAISSOLE, G. CONSTANTINIDES, D. THOMAS.*Formalizing loop-carried dependencies in Coq for high-level synthesis*, April 2019, FCCM 2019 - 27th IEEE Symposium on Field-Programmable Custom Computing Machines, Poster, https://hal.archives-ouvertes.fr/hal-02391954

[45] F. STEINBERG, H. THIES.*Some formal proofs of isomorphy and discontinuity*, March 2019, MLA 2019 - Third Workshop on Mathematical Logic and its Applications, https://hal.inria.fr/hal-02019174

[46] F. STEINBERG, L. THÉRY, H. THIES.*Quantitative continuity and computable analysis in Coq*, April 2019, working paper or preprint, https://hal.inria.fr/hal-02088293

References in notes

[47] A. AYAD, C. MARCHÉ.*Multi-Prover Verification of Floating-Point Programs*, in "Fifth International Joint Conference on Automated Reasoning", Edinburgh, Scotland, J. GIESL, R. HÄHNLE (editors), Lecture Notes in Artificial Intelligence, Springer, July 2010, vol. 6173, p. 127–141, http://hal.inria.fr/inria-00534333

[48] C. BARRETT, C. L. CONWAY, M. DETERS, L. HADAREAN, D. JOVANOVIĆ, T. KING, A. REYNOLDS, C. TINELLI.*CVC4*, in "Proceedings of the 23rd international conference on Computer aided verification", Berlin, Heidelberg, CAV'11, Springer-Verlag, 2011, p. 171–177

[49] P. BAUDIN, J.-C. FILLIÂTRE, C. MARCHÉ, B. MONATE, Y. MOY, V. PREVOSTO.*ACSL: ANSI/ISO C Specification Language, version 1.4*, 2009

[50] P. BEHM, P. BENOIT, A. FAIVRE, J.-M. MEYNADIER.*METEOR : A successful application of B in a large project*, in "Proceedings of FM'99: World Congress on Formal Methods", J. M. WING, J. WOODCOCK, J. DAVIES (editors), Lecture Notes in Computer Science (Springer-Verlag), Springer Verlag, September 1999, p. 369–387

[51] S. BOLDO, F. CLÉMENT, J.-C. FILLIÂTRE, M. MAYERO, G. MELQUIOND, P. WEIS.*Formal Proof of a Wave Equation Resolution Scheme: the Method Error*, in "Proceedings of the First Interactive Theorem Proving Conference", Edinburgh, Scotland, M. KAUFMANN, L. C. PAULSON (editors), LNCS, Springer, July 2010, vol. 6172, p. 147–162, http://hal.inria.fr/inria-00450789/

[52] S. BOLDO, F. CLÉMENT, J.-C. FILLIÂTRE, M. MAYERO, G. MELQUIOND, P. WEIS.*Wave Equation Numerical Resolution: a Comprehensive Mechanized Proof of a C Program*, in "Journal of Automated Reasoning", April 2013, vol. 50, n^o 4, p. 423–456, http://hal.inria.fr/hal-00649240/en/

[53] S. BOLDO, J.-C. FILLIÂTRE, G. MELQUIOND.*Combining Coq and Gappa for Certifying Floating-Point Programs*, in "16th Symposium on the Integration of Symbolic Computation and Mechanised Reasoning", Grand Bend, Canada, Lecture Notes in Artificial Intelligence, Springer, July 2009, vol. 5625, p. 59–74

[54] S. BOLDO, C. MARCHÉ.*Formal verification of numerical programs: from C annotated programs to mechanical proofs*, in "Mathematics in Computer Science", 2011, vol. 5, p. 377–393, http://hal.inria.fr/hal-00777605

[55] S. BOLDO, T. M. T. NGUYEN.*Proofs of numerical programs when the compiler optimizes*, in "Innovations in Systems and Software Engineering", 2011, vol. 7, p. 151–160, http://hal.inria.fr/hal-00777639

[56] L. BURDY, Y. CHEON, D. R. COK, M. D. ERNST, J. R. KINIRY, G. T. LEAVENS, K. R. M. LEINO, E. POLL.*An overview of JML tools and applications*, in "International Journal on Software Tools for Technology Transfer (STTT)", June 2005, vol. 7, n^o 3, p. 212–232

[57] M. CLOCHARD.*Méthodes et outils pour la spécification et la preuve de propriétés difficiles de programmes séquentiels*, Université Paris-Saclay, March 2018, https://tel.archives-ouvertes.fr/tel-01787689

[58] S. CONCHON, A. COQUEREAU, M. IGUERNLALA, A. MEBSOUT.*Alt-Ergo 2.2*, in "SMT Workshop: International Workshop on Satisfiability Modulo Theories", Oxford, United Kingdom, July 2018, https://hal.inria.fr/hal-01960203

[59] C. FUMEX, C. MARCHÉ, Y. MOY.*Automating the Verification of Floating-Point Programs*, in "Verified Software: Theories, Tools, and Experiments. Revised Selected Papers Presented at the 9th International Conference VSTTE", Heidelberg, Germany, A. PASKEVICH, T. WIES (editors), Lecture Notes in Computer Science, Springer, December 2017, n^o 10712, https://hal.inria.fr/hal-01534533/

[60] N. JEANNEROD, C. MARCHÉ, R. TREINEN.*A Formally Verified Interpreter for a Shell-like Programming Language*, in "Verified Software: Theories, Tools, and Experiments. Revised Selected Papers Presented at the 9th International Conference VSTTE", Heidelberg, Germany, A. PASKEVICH, T. WIES (editors), Lecture Notes in Computer Science, Springer, December 2017, n^o 10712, https://hal.inria.fr/hal-01534747v1

[61] N. JEANNEROD, R. TREINEN.*Deciding the First-Order Theory of an Algebra of Feature Trees with Updates*, in "9th International Joint Conference on Automated Reasoning", Oxford, United Kingdom, July 2018, https://hal.archives-ouvertes.fr/hal-01807474

[62] G. KLEIN, J. ANDRONICK, K. ELPHINSTONE, G. HEISER, D. COCK, P. DERRIN, D. ELKADUWE, K. ENGELHARDT, R. KOLANSKI, M. NORRISH, T. SEWELL, H. TUCH, S. WINWOOD.*seL4: Formal verification of an OS kernel*, in "Communications of the ACM", June 2010, vol. 53, n^o 6, p. 107–115

[63] X. LEROY.*A formally verified compiler back-end*, in "Journal of Automated Reasoning", 2009, vol. 43, n^o 4, p. 363–446, http://hal.inria.fr/inria-00360768/en/

[64] A. MAHBOUBI, G. MELQUIOND, T. SIBUT-PINOTE.*Formally Verified Approximations of Definite Integrals*, in "Proceedings of the 7th International Conference on Interactive Theorem Proving", Nancy, France, J. C. BLANCHETTE, S. MERZ (editors), Lecture Notes in Computer Science, August 2016, vol. 9807, https://hal.inria.fr/hal-01289616

[65] J.-M. MULLER, N. BRUNIE, F. DE DINECHIN, C.-P. JEANNEROD, M. JOLDES, V. LEFÈVRE, G. MELQUIOND, N. REVOL, S. TORRES.*Handbook of Floating-point Arithmetic (2nd edition)*, Birkhäuser Basel, July 2018, https://hal.inria.fr/hal-01766584

[66] T. M. T. NGUYEN, C. MARCHÉ.*Hardware-Dependent Proofs of Numerical Programs*, in "Certified Programs and Proofs", J.-P. JOUANNAUD, Z. SHAO (editors), Lecture Notes in Computer Science, Springer, December 2011, p. 314–329, http://hal.inria.fr/hal-00772508

[67] R. RIEU-HELFT.*Un mécanisme d'extraction vers C pour Why3*, in "Vingt-neuvièmes Journées Francophones des Langages Applicatifs", Banyuls-sur-mer, France, S. BOLDO, N. MAGAUD (editors), January 2018, https://hal.inria.fr/hal-01653153

[68] F. DE DINECHIN, C. LAUTER, G. MELQUIOND.*Certifying the floating-point implementation of an elementary function using Gappa*, in "IEEE Transactions on Computers", 2011, vol. 60, n^o 2, p. 242–253, http://hal.inria.fr/inria-00533968/en/

[69] S. DE GOUW, J. ROT, F. S. DE BOER, R. BUBEL, R. HÄHNLE.*OpenJDK's Java.utils.Collection.sort() Is Broken: The Good, the Bad and the Worst Case*, D. KROENING, C. S. PĂSĂREANU (editors), Springer International Publishing, Cham, 2015, p. 273–289

[70] L. DE MOURA, N. BJØRNER.*Z3, An Efficient SMT Solver*, in "TACAS", Lecture Notes in Computer Science, Springer, 2008, vol. 4963, p. 337–340

Project-Team TRIBE

inTeRnet BEyond the usual

RESEARCH CENTER
Saclay - Île-de-France

THEME
Networks and Telecommunications

Table of contents

Project-Team TRIBE

Creation of the Project-Team: 2019 June 01

Keywords:

Computer Science and Digital Science:

A1.2.3. - Routing
A1.2.5. - Internet of things
A1.2.6. - Sensor networks
A1.2.7. - Cyber-physical systems
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A1.4. - Ubiquitous Systems
A2.6.1. - Operating systems
A3.1.1. - Modeling, representation
A3.2.2. - Knowledge extraction, cleaning
A3.3.2. - Data mining
A7.1. - Algorithms
A7.1.3. - Graph algorithms
A8.6. - Information theory
A9.2. - Machine learning

Other Research Topics and Application Domains:

B4.4. - Energy delivery
B4.4.1. - Smart grids
B6.3.2. - Network protocols
B6.3.3. - Network Management
B6.4. - Internet of things
B6.6. - Embedded systems
B7.2.1. - Smart vehicles
B8.1.2. - Sensor networks for smart buildings
B8.2. - Connected city
B9.5.1. - Computer science

1. Team, Visitors, External Collaborators

Research Scientists

Aline Carneiro Viana [Team leader, Inria, Researcher, from Jun 2019, HDR]
Cedric Adjih [Inria, Researcher, from Jun 2019]
Emmanuel Baccelli [Inria, Researcher, from Jun 2019, HDR]
Philippe Jacquet [Inria, Senior Researcher, from Nov 2019, HDR]

Faculty Member

Nadjib Achir [Univ Paris-Nord, Associate Professor, from Sep 2019, HDR]

Technical Staff

Oumaima Attia [Inria, Engineer, from Mar 2019]

François-Xavier Molina [Inria, Engineer, from Mar 2019]

PhD Students

Licia Amichi [Inria, PhD Student, from Oct 2018]
Rafael Costa [CAPES, PhD Student, from May 2017, co-tutelle]
Anne Josiane Kouam Djuigne [Inria, PhD Student, from Nov 2019]
Douglas Do Couto Teixeira [CAPES, PhD Student, from Apr 2018, co-tutelle]
Diego Madariaga Roman [CONICYT, PhD Student, from Jan 2019, co-tutelle]
Hirah Malik [Inria, PhD Student, from Jun 2019]
Iman Hmedoush [Inria, from Oct 2018, co-advised with EVA]

Administrative Assistant

Laurence Fontana [Inria, Administrative Assistant, from Jun 2019]

2. Overall Objectives

2.1. Vision and approach

TRiBE stands for "Internet Beyond the Usual" and belongs to the Inria theme "Networks and Telecommunications" as well as contributes to the "Challenge no 11: Toward a trustworthy Internet of Everything" of the strategic plan of Inria. *Building on an approach combining protocol design, data analytics, and experimental research, the research contributions of TRiBE aims at contributing to the design of smart, unified, and tactful Internet edge networks, skilled for answering application, services, or end-users' purposes.*

All the emerging IoT specificities and requirements (i.e., heterogeneity of devices and services, densification, traffic growth, ubiquitous cyber-physical context, etc) bring new demands and consequently, new scientific and technological challenges to the edge of the Internet. In this context, our conviction is that the success of the Internet of Things is rooted in the **network designing choices** involving its devices and related protocols/services as well as the edge-core network communication loop.

Toward this belief, the research of the team will be organized around three research directions: *(1) technologies for accomodating low-end IoT devices; (2) technologies for leveraging high-end IoT devices' advents; and (3) technologies for edge-core network interaction.* With those three research directions, the team place its efforts in the three main elements composing the ecosystem of IoT devices: (1) the device itself, (2) their usability, and (3) their network context.

More specifically, the first element tackles the optimization, simplicity, and unification requirements imposed by the heterogeneity and low capabilities of low-end IoT devices. This brings the necessity to deal with hardware and software specificity of devices, while adapting designing choices and simplifying deployment. The second element focus on issues related to "how" and "for what" IoT devices are used. This also brings the human element into play, which dynamics are shaping the way their mobile devices are interacting with the edge of the Internet and consequently, are requesting and consuming network resources and services. Finally, the third element closes the *network→usability→device→network* loop by bringing solutions supporting functions and communication between IoT devices and the core of the Internet.

2.2. New challenges

The Internet has steadily evolved over the past decades from a small homogeneous to a gigantic Internet of Things (IoT) interconnecting an extremely wide variety of machines (e.g., PCs, smart-phones, sensors/actuators, smart appliances, smart vehicles), and providing an extremely wide variety of services. Globally, devices and connections are growing faster than both the population and Internet users, as foreseen by Cisco. Forecasts mention an IoT market that will attain a compound annual growth rate of 28.5% from 2016 to 2020 as well as an installed base of IoT devices over 75.4B devices by 2025. Added to these statistics is the fact that global mobile data traffic will grow nearly twice as fast as fixed IP traffic from 2017 to 2022: Smartphones account for most of this growth.

Hence, the edge of this network now consists in a dense deployment of machines ranging from PCs to smartphones, from sensors/actuators to smart appliances, and from smart vehicles to diverse kinds of robots. As a consequence, humans are immersed in a highly connected and ubiquitous cyber-physical context, and as end-users of the network and its numerous services, their satisfaction has become the main focus.

In this context, the IoT is simultaneously used as a tool to gather more data, and as a mean to automate more advanced control. Some businesses and institutions aim to gather more data to better understand their customers, so as to improve services. Others efforts aim to further immerse their customers into a flourishing, integrated cyber-physical environment, which can automatically and optimally adapts to their needs. All these emerging IoT-related opportunities bring new requirements and consequently, new scientific and technological challenges to the edge of the Internet.

First, the densified deployment of heterogeneous **low-end IoT devices** (e.g. sensors, actuators, etc.) at the edge of the Internet requires to deal with (1) the accommodation of machines with extremely limited capabilities, with a primary focus on low power requirements while (2) allowing their seamless integration in interoperable systems (often using IP as a common factor).

Second, today's pervasiveness of **high-end IoT devices** (e.g. smart handheld devices) distribute increasing capabilities (i.e., processing, storage, connectivity) at the edge of the network, and make our real life and virtual activities seamlessly merged together. In this domain, we need a better understanding of: (1) when, where, and for what the high-end IoT devices are used, (2) how the uses vary among individuals, and (3) how social norms and structure dictating individuals' behavior influence the way they interact with network services and demand resources.

The research contributions of the team aims thus, at dealing with such requirements and challenges brought to the edge of the Intenet. One should design adequate algorithms and communication mechanisms for addressing such challenges as well as for leveraging the new technological opportunities brought by the Internet of Things.

3. Research Program

3.1. Research program

Following up on the effort initiated by the team members during the last few years and building on an approach combining protocol design, data analytics, and experimental research, we propose a research program organized around three closely related objectives that are briefly described in the following.

- **Technologies for accommodating low-end IoT devices:** The IoT is expected to gradually connect billions of low-end devices to the Internet, and thereby drastically increase communication without human source or destination. Low-end IoT devices differ starkly from high-end IoT devices in terms of resources such as energy, memory, and computational power. Projections show this divide will not fundamentally change in the future and that IoT should ultimately interconnect a dense population of devices as tiny as dust particles, feeding off ambient power sources (energy harvesting). These characteristics constrain the software and communication protocols running on low-end IoT devices: they are neither able to run a common software platform such as Linux (or its derivatives), nor the standard protocol stack based on TCP/IP. Solutions for low-end IoT devices require thus: **(i)optimized communication protocols** taking into account radio technology evolution and devices constrained requirements; **(ii)tailored software platforms** providing high level programming, modular software updates as well as advanced support for new security and energy concentration features; **(iii)unification of technologies** for low-end IoT, which is too fragmented at the moment, guaranteeing integration with core or other edge networks.
- **Technologies for leveraging high-end IoT devices' advents:** High-end IoT devices are one of the most important instances of the connected devices supporting a noteworthy shift towards mobile Internet access. As our lives become more dependent on pervasive connectivity, our social patterns (as human being in the Internet era) are nowadays being reflected from our real life onto the

virtual binary world. This gives birth to two tendencies. From one side, edge networks can now be utilized as mirrors to reflect the inherent human dynamics, their context, and interests thanks to their well organized recording and almost ubiquitous coverage. From the other side, social norms and structure dictating human behavior (e.g., interactions, mobility, interest, cultural patterns) are now directly influencing the way individuals interact with the network services and demand resources or content. In particular, we observe the particularities present in human dynamics *shape the way (i.e., where, when, how, or what) resources, services, and infrastructures are used at the edge of the Internet*. Hence, we claim a need to digitally study high-end IoT devices' end-users behaviors and to leverage this understanding in networking solutions' design, so as to optimize network exploitation. This suggests the **integration of the heterogeneity and uncertainty of behaviors in designed networking solutions**. For this, *useful knowledge* allowing the understanding of behaviors and context of users has to be *extracted and delivered out* of large masses of data. Such knowledge has to be then *integrated in current design practices*. This brings the idea of a more *tactful networking design practice* where the network is assigned with the human like capability of observation, interpretation, and reaction to daily life features and entities involving high-end IoT devices. Research activities here include: **(i) the quest for meaningful data**, which includes the integration of data from different sources, the need for scaling up data analysis, the usage and analysis of fine-grained datasets, or still, the completion of sparse and coarse grained datasets; **(ii) expanding edge networks' usage understanding**, which concerns analysis on how and when contextual information impact network usage, fine-grained analysis of short-term mobility of individuals, or the identification of patterns of behavior and novelty-seeking of individuals; **(iii) human-driven prediction models**, extensible to context awareness and adapted to individuals preferences in terms of novelty, difersity, or routines.

- **Articulating the IoT edge with the core of the network:** The edge is the interface between the IoT devices and the core network: some of the challenges encountered by IoT devices have their continuity at the edge of the network inside the gateway (i.e., interoperability, heterogeneity and mobility support). Besides, the edge should be able to support intermediary functions between devices and the rest of the core (e.g., the cloud). This includes: **(i) proxying functionality**, facilitating connections between devices and the Internet; **(ii) machine learning enhanced IoT solutions**, designed to improve performance of advanced IoT networked systems (e.g., through methods such as supervised, unsupervised or reinforcement learning) at adapted levels of the protocol stack (e.g., for multiple access, coding, choices); **(iii) IoT data contextualization**, so collection of meaningful IoT data (i.e., right data collected at the right time) can be earlier determined closer to the data source; **(iv) intermediary computation** through fog or Mobile Edge Computing (MEC) models, where IoT devices can obtain computing, data storage, and communication means with lower latency in a decentralized way; or **(v) security of end-to-end IoT software supply-chain**, including remote management and over-the-air updates.

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Together with his co-authors, Aline Carneiro Viana was awarded: (1) the **best poster award** at the main conference on the scientific analysis of mobile phone datasets (NetMob) in Oxford, UK, Jul. 2019, for the poster on **Complete Trajectory Reconstruction from Sparse Mobile Phone Data** (collaboration with G. Chen, M. Fiore, and C. Sarraute); (2) the **top-six best paper award** at the 27th International Conference on Advances in Geographic Information Systems 2019 (ACM SIGSPATIAL), in Chicago, USA, Nov. 2019, for the paper on **Deciphering Predictability Limits in Human Mobility**.

4.1.2. RIOT Summit 2019

We successfully organized in September 2019 the fourth RIOT Summit, in Helsinki. The RIOT Summit 2019 gathered ~100 enthusiastic industrial participants, makers and academics involved in RIOT. Highlights included a keynote from IoT expert and former IETF general chair Jari Arkko (Ericsson Research), and a new car-sharing product using RIOT announced by Continental (and now deployed on thousands of vehicles). Aside of big companies and academics, a number of SMEs and startups from various places in Europe gave talks on aspects of IoT communication, use cases IoT hardware, IoT open source community aspects and concepts for future IoT software and networks, as well as hands-on sessions and tutorials. See: http://summit.riot-os.org.

4.1.3. Associated team - EMBRACE

2019 was the third and last year of the EMBRACE Associated team. The EMBRACE (lEveraging huMan Behavior for Resource AlloCation and services orchestration modEls) team was composed by members of the INFINE and by three Brazilian teams from three different Brazilian Universities. The EMBRACE project addressed the topic of designing efficient solutions for 5G networks taking into account human behavior, uncertainty, and heterogeneity of networking resources. A proposal requesting the extension of the project was submitted in Nov. 2019.

More information is available here: https://team.inria.fr/embrace/.

4.1.4. IETF Hackathons

Concerning Internet Standardization, we contributed to all three IETF Hackathons in 2019. In particular, Oumaima Attia and Cedric Adjih were some major contributors (with many others including Vincent Roca, EPI Privatics) in the NWCRG Hackathon which allowed to release in first prototype of SWIF-codec, a sliding-window forward-error correction codec, see: **https://github.com/irtf-nwcrg/swif-codec**. Cedric Adjih is also a major contributor to the LPWAN Hackathon on the SCHC protocol (IPv6 compression for IoT networks), which resulted this year in code at **https://github.com/openschc/openschc**, a working prototype of the protocol.

5. New Software and Platforms

5.1. RIOT

KEYWORDS: Internet of things - Operating system - Sensors - Iot - Wireless Sensor Networks - Internet protocols

SCIENTIFIC DESCRIPTION: While requiring as low as 1,5kB of RAM and 5kB or ROM, RIOT offers real time and energy efficiency capabilities, as well as a single API (partially POSIX compliant) across heterogeneous 8-bit, 16-bit and 32-bit low-hardware. This API is developer-friendly in that it enables multi-threading, standard C and C++ application programming and the use of standard debugging tools (which was not possible so far for embedded programming). On top of this, RIOT includes several network stacks, such as a standard IPv6/6LoWPAN stack and a information-centric network stack (based on CCN).

FUNCTIONAL DESCRIPTION: RIOT is an Open Source operating system that provides standard protocols for embedded systems. RIOT allows, for example, the development of applications that collect sensor data and transmit it to a central node (e.g. a server). This data can then be used for smart energy management for instance.

RIOT is specially designed for embedded systems, which are strongly constrained in memory and energy. Further, RIOT can easily be ported to different hardware devices and follows the latest evolution of IP standards.

RIOT applications can readily be tested in the FIT IoT-Lab, which provides a large-scale infrastructure facility with 3000 nodes for testing remotely small wireless devices.

- Participants: Emmanuel Baccelli and Oliver Hahm
- Partners: Freie Universität Berlin - University of Hamburg
- Contact: Emmanuel Baccelli
- URL: http://www.riot-os.org

5.2. openschc

KEYWORDS: Internet of things - Internet protocols - Low-Power WIreless

FUNCTIONAL DESCRIPTION: OpenSCHC is a OpenSource Implementation of SCHC (Static Context Header Compression) currently being standardized by the LPWAN Working Group at the IETF. Oversimplifying, this is essentially IPv6 compression and fragmentation intended for low datarate, long range IoT networks.

The short/medium term goal is to organize SCHC Hackathons that occur at the IETF meetings.

The long term wish is to have a stable, open-source, reference codebase for the SCHC protocol (in Python).

The authors of OpenSCHC are listed here: https://github.com/openschc/openschc/blob/master/AUTHORS.txt

- Partner: OpenSCHC Authors
- Contact: Cédric Adjih
- URL: https://github.com/openschc/openschc

5.3. Gardinet

KEYWORD: Distributed networks

FUNCTIONAL DESCRIPTION: Gardinet (previously DragonNet) is a generic framework for network coding in wireless networks. It is a initially result of the GETRF project of the Hipercom2 team.

It is based on intra-flow coding where the source divides the flow in a sequence of payloads of equal size (padding may be used). The design keys of DragonNet are simplicity and universality, DragonNet does not use explicit or implicit knowledge about the topology (such as the direction or distance to the source, the loss rate of the links, ...). Hence, it is perfectly suited to the most dynamic wireless networks. The protocol is distributed and requires minimal coordination. DragonNet architecture is modular, it is based on 5 building blocks (LIB, SIG, Protocol, SEW and DRAGON). Each block is almost independent. This makes DragonNet generic and hence adaptable to many application scenarios. DragonNet derives from a prior protocol called DRAGONCAST. Indeed, DragonNet shares the same principles and theoretical overview of DRAGONCAST. It enriches DRAGONCAST by the information base and signaling required to perform broadcast in wireless networks and in wireless sensor networks in particular.

- Participants: Antonia Masucci, Cédric Adjih, Hana Baccouch and Ichrak Amdouni
- Contact: Cédric Adjih
- URL: http://gitlab.inria.fr/gardinet

6. New Results

6.1. Human Mobility completion of Sparse Call Detail Records

Participants: Guangshuo Chen [Inria], Aline Carneiro Viana, Marco Fiore [CNR], Carlos Sarraute [Gran-Data].

Mobile phone data are a popular source of positioning information in many recent studies that have largely improved our understanding of human mobility. These data consist of time-stamped and geo-referenced communication events recorded by network operators, on a per-subscriber basis. They allow for unprecedented tracking of populations of millions of individuals over long time periods that span months. Nevertheless, due to the uneven processes that govern mobile communications, the sampling of user locations provided by mobile phone data tends to be sparse and irregular in time, leading to substantial gaps in the resulting trajectory information. In this work, we illustrate the severity of the problem through an empirical study of a large-scale Call Detail Records (CDR) dataset. We then propose two novel and effective techniques to reduce temporal sparsity in CDR that outperform existing ones. the fist technique performs completion (1) at nightime by identifying temporal home boundary and (2) at daytime by inferring temporal boundaries of users, i.e., the time span of the cell position associated with each communication activity. The second technique, named Context-enhanced Trajectory Reconstruction, complete individual CDR-based trajectories that hinges on tensor factorization as a core method by leveraging regularity in human movement patterns.

Our approach lets us revisit seminal works in the light of complete mobility data, unveiling potential biases that incomplete trajectories obtained from legacy CDR induce on key results about human mobility laws, trajectory uniqueness, and movement predictability. In addition, the CTR solution infers missing locations with a median displacement within two network cells from the actual position of the user, on a hourly basis and even when as little as 1% of her original mobility is known.

These works have been published at two journals: EPJ Data Science in 2019 and at Computer Communication Elsevier in 2018.

6.2. Adaptive sampling frequency of human mobility

Participants: Panagiota Katsikouli [AGORA], Aline Carneiro Viana, Marco Fiore [CNR], Diego Madariaga.

In recent years, mobile device tracking technologies based on various positioning systems have made location data collection a ubiquitous practice. Applications running on smartphones record location samples at different frequencies for varied purposes.The frequency at which location samples are recorded is usually pre-defined and fixed but can differ across applications; this naturally results in big location datasets of various resolutions. What is more, continuous recording of locations results usually in redundant information, as humans tend to spend significant amount of their time either static or in routine trips, and drains the battery of the recording device.

In this work, we aim at answering the question *"at what frequency should one sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*. Our first analyses on fine-grained GPS trajectories from users around the world unveil *(i)* seemingly universal spectral properties of human mobility, and *(ii)* a linear scaling law of the localization error with respect to the sampling interval. Such results were published at a paper at IEEE Globecom 2017.

Building on these results, we challenge the idea of a fixed sampling frequency and present a lightweight mobility aware adaptive location sampling mechanism. This is an on-going work with Panagiota Katsikouli, who spent 5 months in our team working as an internship in 2017, and Diego Madariaga who spent 3 months in 2018 in our team working as an internship and has started a PhD in co-tutelle with Aline C. Viana and Javier Bustos (NIC/Univ. of Chile).

Our mechanism can serve as a standalone application for *adaptive location sampling*, or as complimentary tool alongside auxiliary sensors (such as accelerometer and gyroscope). In this work, we implemented our mechanism as an application for mobile devices and tested it on mobile users worldwide. Our experiments show that our method adjusts the sampling frequency to the mobility habits of the tracked users, it reliably tracks a mobile user incurring acceptable approximation errors and significantly reduces the energy consumption of the mobile device.

A journal paper is being prepared for submission.

6.3. Inference of human personality from mobile phones datasets

Participants: Adriano Di Luzio [Sapienza U. di Rome], Aline Carneiro Viana, Julinda Stefa [Sapienza U. di Rome], Katia Jaffres-Runser [U. of Toulouse], Alessandro Mei [Sapienza U. di Rome].

Related to human behavioral studies, personality prediction research has enjoyed a strong resurgence over the past decade. Due to the recognition that personality is predictive of a wide range of behavioral and social outcomes, the human migration to the digital environment renders also possible to base prediction of individual personality traits on digital records (i.e., datasets) mirroring human behaviors. In psychology, one of the most commonly used personality model is the Big5, based on five crucial traits and commonly abbreviated as OCEAN: Openness (O), Conscientiousness (C), Extroversion (E), Agreeableness (A), and Neuroticism (N). They are relatively stable over time, differ across individuals, and, most importantly, guide our emotions and our reactions to life circumstances. It is so for social and work situations, and even for things as simple as the way we use our smartphone. For instance, a person that is curious and open to new experiences will tend to look continuously for new places to visit and thrills to experience.

This work brings the deepest investigation in the literature on the prediction of human personality (*i.e.,* captured by the Big5 traits) from smartphone data describing daily routines and habits of individuals. This work shows that human personality can be accurately predicted by looking at the data generated by our smartphones. GPS location, calls, battery usage and charging, networking context like bluetooth devices and WiFi access points in proximity, and more give enough information about individual habits, reactions, and idiosyncrasies to make it possible to infer the psychological traits of the user. We demonstrate this by using machine learning techniques on a dataset of 55 volunteers who took a psychological test and allowed continuous collection of data from their smartphones for a time span of up to three years. Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism (the so called Big5 personality traits) can be predicted with good accuracy even by using just a handful of features. The possible applications of our findings go from network optimization, to personal advertising, and to the detection of mental instability and social hardship in cities and neighborhoods. We also discuss the ethical concerns of our work, its privacy implications, and ways to tradeoff privacy and benefits.

A paper describing this work is under submission at ACM Transactions on Data Science (TDS) , but a technical report is also registered under the name hal-01954733.

6.4. Data offloading decision via mobile crowdsensing

Participants: Emanuel Lima [U. of Porto], Aline Carneiro Viana, Ana Aguiar [U. of Porto], Paulo Carvalho [Univ. Do Minho].

According to Cisco forecasts, mobile data traffic will grow at a compound annual growth rate of 47 % from 2016 to 2021 with smartphones surpassing four-fifths of mobile data traffic. It is known that mobile network operators are struggling to keep up with such traffic demand, and part of the solution is to offload communications to WiFi networks. Mobile data offloading systems can assist mobile devices in the decision making of when and what to offload to WiFi networks. However, due to the limited coverage of a WiFi AP, the expected offloading performance of such a system is linked with the users mobility. Unveiling and understanding human mobility patterns is a crucial issue in supporting decisions and prediction activities for mobile data offloading.

Several studies on the analysis of human mobility patterns have been carried out focusing on the identification and characterization of important locations in users' life in general. We extended these works by studying human mobility from the perspective of mobile data offloading. In our study, offloading zones are identified and characterized from individual GPS trajectories when small offloading time windows are considered. The characterization is performed in terms availability, sojourn, transition time; type and spatial characteristics. We then evaluate the offloading opportunities provided to users while they are travelling in terms of availability, time window to offload and offloading delay. We also study the mobility predictability in an offloading scenario through the theoretical and practical evaluation of several mobility predictors. The results show that (i) attending to users mobility, ten seconds is the minimum offloading time window that can be considered; (ii) offloading predictive methods can have variable performance according to the period of the day; and (iii) per-user opportunistic decision models can determine offloading system design and performance.

This work was published at ACM CHANTS 2018 and its extension will be submitted to WoWMON 2020. This is an on-going work with the the PhD Emanuel Lima (one of my co-supervision), who spent 4 months as an intern in our team in 2018, and his advisors.

6.5. Identifying how places impact each other by means of user mobility

Participants: Lucas Santos de Oliveira [EMBRACE], Pedro Olmo Stancioli [Federal U. of Minas Gerais], Aline Carneiro Viana.

The way in which city neighborhoods become popular and how people trajectory impacts the number of visitation is a fundamental area of study in traditional urban studies literature. Many works address this problem by means of user mobility prediction and POI recommendation. In a different approach, other works address the human mobility in terms of social influence which refers to the case when individuals change their behaviors persuaded by others. Nevertheless, fewer works measure influence of POI based on human mobility data.

Different from previous literature, in this work, we are interested in understanding how the neighborhood POI affect each other by means of human mobility using location-based social networks (LBSNs) data source. Key location identification in cities is a central in human mobility investigation as well as for societal problem comprehension. In this context, we propose a methodology to quantify the power of point-of-interests (POIs) in their vicinity, in terms of impact and independence – the first work in the literature (to the best of our knowledge). Different from literature, we consider the flow of people in our analysis, instead of the number of neighbor POIs or their structural locations in the city. Thus, we first modeled POI's visits using the multiflow graph model where each POI is a node and the transitions of users among POIs are a weighted direct edge. Using this multiflow graph model, we compute the attract, support and independence powers. The attract power and support power measure how many visits a POI gather from and disseminate over its neighborhood, respectively. Moreover, the independence power captures the capacity of POI to receive visitors independently from other POIs. Using a dataset describing the mobility of individuals in the Dartmouth College campus, we identify a slight dependence among buildings as well as the tendency of people to be mostly stationary in few buildings with short transit periods among them.

This work was published in ACM MobiWac 2019 [14] and an extended version is being prepared. Lucas is doing an internship in our team from Nov. 2019 to Jan. 2020.

6.6. Infering friends in the crowd in Device-to-Device communication

Participants: Rafael Lima Da Costa [CAPES], Aline Carneiro Viana, Leobino Sampaio [Federal U. of Bahia], Artur Ziviani [LNCC].

The next generation of mobile phone networks (5G) will have to deal with spectrum bottleneck and other major challenges to serve more users with high-demanding requirements. Among those are higher scalability and data rates, lower latencies and energy consumption plus reliable ubiquitous connectivity. Thus, there is a need for a better spectrum reuse and data offloading in cellular networks while meeting user expectations. According to literature, one of the 10 key enabling technologies for 5G is device-to-device (D2D) communications, an approach based on direct user involvement. Nowadays, mobile devices are attached to human daily life activities, and therefore communication architectures using context and human behavior information are promising for the future. User-centric communication arose as an alternative to increase capillarity and to offload data traffic in cellular networks through opportunistic connections among users. Although having the user as main concern, solutions in the user-centric communication/networking area still do not see the user as an individual, but as a network active element. Hence, these solutions tend to only consider user features that can be measured from the network point of view, ignoring the ones that are intrinsic from human activity (e.g., daily routines, personality traits, etc).

In this work, we first introduce the Tactful Networking paradigm, whose goal is to add perceptive senses to the network, by assigning it with human-like capabilities of observation, interpretation, and reaction to daily-life features and involved entities. To achieve this, knowledge extracted from human inherent behavior (routines, personality, interactions, preferences, among others) is leveraged, empowering user-needs learning and prediction to improve QoE while respecting privacy. We survey the area, propose a framework for enhancing human raw data to assist networking solutions and discuss the tactful networking impact through representative examples. Finally, we outline challenges and opportunities for future research. This tutorial paper is under submittion to ACM Computing and Surveys and a technical report is registered as hal-01675445.

Besides, we investigate how human-aspects and behavior can be useful to leverage future device-to-device communication. We have designed a strategy to select next-hops in a D2D communication that will be human-aware: i.e., that will consider not only available physical resources at the mobile device of a wireless neighbor, her mobility features and restrictions but also any information allowing to infer how much sharing willing she is. Such forwarders nodes will be then used at the offloading of content data through Device-to-Device (D2D) communication, from devices to the closest Mobile Edge Computing infrastructure, transforming mobile phone neighbors in service providers. The selection of next hops based on mobility behavior, resource capability as well as collaboration constitute the novelty we plan to exploit. A conference paper is under preparation and a Brazilian paper under submission to SBRC 2020.

6.7. Deciphering Predictability Limits in Human Mobility

Participants: Douglas Do Couto Teixiera, Aline Carneiro Viana, Jussara Almeida [Federal U. of Minas Gerais], Mario S. Alvim [Federal U. of Minas Gerais].

Human mobility has been studied from different perspectives. One approach addresses predictability, deriving theoretical limits on the accuracy that any prediction model can achieve in a given dataset. Measuring the predictability of any phenomenon is a very useful, but hard task, and especially so in the case of human behavior. Such complexity is due to the uncertain and heterogeneous behavior of humans, as well as to the variability of parameters influencing such behavior. Predictability is concerned with the maximum theoretical accuracy that an ideal prediction model could achieve in a scenario expressed by a given dataset. As such, unlike particular comparisons of alternative prediction models on different datasets, it does not depend on a specific prediction strategy but rather on human behavior, as captured by the available data. Besides, it does not rely on the tuning of a multitude of sensible parameters, providing instead a parameter-free view of how predictable human mobility can be (as expressed in the data).

This approach focuses on the inherent nature and fundamental patterns of human behavior captured in the dataset, filtering out factors that depend on the specificities of the prediction method adopted. In this work, we revisit the state-of-the-art method for estimating the predictability of a person's mobility, which, despite being widely adopted, suffers from low interpretability and disregards external factors that have been suggested to improve predictability estimation, notably the use of contextual information (e.g., weather, day of the week, and time of the day). We propose a new measure, *regularity*, which together with *stationarity*, helps us understand what makes a person's mobility trajectory more or less predictable, as captured by Song et al.'s technique. We show that these two simple measures are complementary and jointly are able to explain most of the variation in Song et al.'s predictability. As such, we here use them as proxies of that technique to analyze how one's mobility predictability varies.

Additionally, we investigate strategies to incorporate different types of contextual information into predictability estimates. In particular, we were the first to quantify the impact of different types of contextual information on predictability in human mobility, for different prediction tasks and datasets. Our results show that, for the next place prediction problem, the use of contextual information plays a larger role than one's history of visited locations in estimating their predictability. Finally, we propose and evaluate alternative estimates of predictability which, while being much easier to interpret, provide comparable results to the state-of-the-art. We show that these estimators, while being more interpretable, provide comparable results in terms of predictability.

This paper was published at ACM SIGSPATIAL 2019, a A+-ranked conference in our domain, and was indicated as a top-six best paper candidate. An extended version is being prepared for submission to a journal.

6.8. Identifying and profiling novelty-seeking behavior in human mobility

Participants: Licia Amichi, Aline Carneiro Viana, Mark Corvella [Boston Univ.], Antonio F. Loureiro [Federal U. of Minas Gerais].

The prediction of individuals' dynamics has attracted significant community attention and has implication for many fields: e.g. epidemic spreading, urban planning, recommendation systems. Current prediction models, however, are unable to capture uncertainties in the mobility behavior of individuals, and consequently, suffer from *the inability to predict visits to new places*. This is due to the fact that current models are oblivious to the exploration aspect of human behavior.

Many prediction models have been proposed to forecast individuals trajectories. However, they all show limited bounded predictive performance. Regardless of the applied methods (e.g., Markov chains, Naive Bayes, neural networks), the type of prediction (i.e., next-cell or next place) or the used data sets (e.g., GPS, CDR, surveys), accuracy of prediction never reaches the coveted 100%. The reasons for such limitations in the accuracy are manyfold: the lack of ground truth data, human beings' complex nature and behavior, as well the exploration phenomenon (i.e., visits to never seen before places). In this work, we focus on the exploration problem, which has rarely been tackled in the literature but indeed, represents a real issue. By construction, most prediction models attempt to forecast future locations from the set of known places, which hinders predicting new unseen places and by consequence, reduces the predictive performance.

Thus, when considering the exploration problem, previous studies either did not provide any consideration of the exploration factors of individuals, or divided the population based on properties that are not always consistent, or assumed that all individuals have the same propensity to explore. Our main goal in this work is to understand the exploration phenomenon and answer the following question: *What type of visits characterize the mobility of individuals?* Using newly designed metrics capturing spatiotemporal properties of human mobility – i.e., known/new and recurrent/intermittent visits – our strategy identifies three groups of individuals according to their degree of exploration: scouters, routineers, and regulars. In the future, we plan to deeply investigate the mobility behavior of individuals in each profile and to assign to each individual an *exploration factor* describing her susceptibility to explore.

This work was published at the Student workshop of ACM CONEXT 2019 [9]. An extended version is being prepared for submission to an int. conference.

6.9. How Geo-indistinguishability Affects Utility in Mobility-based Geographic Datasets

Participants: Adriano Di Luzio [Inria], Aline Carneiro Viana, Catuscia Palamidessi [Comete – Inria], Konstantinos Chatzikokolakis [Comete – Inria], Georgi Dikov [Comete – Inria], Julinda Stefa [Sapienza University].

Many of the scientific challenges that we face today deal with improving the quality of our everyday lives. They aim at making the cities around us smarter, more efficient, and more sustainable (e.g., how to schedule public transport during peak hours or what is the most efficient path for waste disposal). All these challenges share a common ground. They rely on datasets gathered from the real world that depict the mobility of hundreds of thousands individuals and picture, with great detail, the whereabouts of their lives—where they live, work, shop for groceries, and hangout with friends. At the same time, however, the collection of personal data also endangers the privacy of the users that to whom these data belong. To protect the privacy of the users, it is necessary to sanitize these datasets before releasing them to the public.

When we sanitize the datasets we trade the accuracy of the information they contain to protect the privacy of their users. The task of this work is to shed light on the effects of the trade-off between privacy and utility in mobility-based geographic datasets. We aim at finding out whether it is possible to protect the privacy of the users in a dataset while, at the same time, maintaining intact the utility of the information that it contains. In particular, we focus on geo-indistinguishability as a privacy-preserving sanitization methodology, and we evaluate its effects on the utility of the Geolife dataset. We test the sanitized dataset in two real world scenarios: (1) Deploying an infrastructure of WiFi hotspots to offload the mobile traffic of users living, working, or commuting in a wide geographic area; (2) Simulating the spreading of a gossip-based epidemic as the outcome of a device-to-device communication protocol. We show the extent to which the current geo-indistinguishability techniques trade privacy for utility in real world applications and we focus on their effects at the levels of the population as a whole and of single individuals.

This paper was published at the LocalRec 2019 workshop, jointly with ACM SIGSPATIAL [12].

6.10. General-purpose Low-power Secure Firmware Updates for Constrained IoT Devices

Participants: Koen Zandberg [Inria / Freie Universität Berlin], Kaspar Schleiser [Inria / Freie Universität Berlin], Francisco Acosta [Inria], Hannes Tschofenig [Arm Ltd., Cambridge, U.K], Emmanuel Baccelli.

While the IoT deployments multiply in a wide variety of verticals, the most IoT devices lack a built-in secure firmware update mechanism. Without such a mechanism, however, critical security vulnerabilities cannot be fixed, and the IoT devices can become a permanent liability, as demonstrated by recent large-scale attacks. In this paper, we survey open standards and open source libraries that provide useful building blocks for secure firmware updates for the constrained IoT devices–by which we mean low-power, microcontroller-based devices such as networked sensors/actuators with a small amount of memory,among other constraints. We design and implement a prototype that leverages these building blocks and assess the security properties of this prototype. We present experimental results including first experiments with SUIT, a new IETF standard for secure IoT firmware updates. We evaluate the performance of our implementation on a variety of commercial off-the-shelf constrained IoT devices. We conclude that it is possible to create a secure, standards-compliant firmware update solution that uses the state-of-the-art security for the IoT devices with less than 32 kB of RAM and 128 kB of flash memory. Moreover, our prototype is general-purpose, in that it works out-of-the-box or with minimal adaptation on 80% of the hardware supported by RIOT (i.e. approximately 100 different types of IoT devices). As such, this work paves the way towards generic and secure low-power IoT firmware updates.

This paper was published in the IEEE journal IEEE Access [8].

6.11. LoRa-MAB: A Flexible Simulator for Decentralized Learning Resource Allocation in IoT Networks

Participants: Duc-Tuyen Ta [LRI and Inria], Kinda Khawam [UVSQ], Samer Lahoud [ESIB], Cédric Adjih, Steven Martin [LRI, Université Paris-Saclay].

LoRaWAN is a media access control (MAC) protocol for wide area networks. It is designed to allow low-powered devices to communicate with Internet-connected applications over long-range wireless connections. The targeted dense deployment will inevitably cause a shortage of radio resources. Hence, autonomous and lightweight radio resource management is crucial to offer ultra-long battery lifetime for LoRa devices. One of the most promising solutions to such a challenge is the use of artificial intelligence. This will enable LoRa devices to use innovative and inherently distributed learning techniques, thus freeing them from draining their limited energy by constantly communicating with a centralized controller.Before proceeding with the deployment of self-managing solutions on top of a LoRaWAN application, it is sensible to conduct simulation-based studies to optimize the design of learning-based algorithms as well as the application under consideration. Unfortunately, a network simulator for such a context is not fully considered or lacks real deployment parameters. In order to address this shortcoming, we have developed an event-based simulator for resource allocation in LoRaWAN. To demonstrate the usefulness of our simulator, extensive simulations were run in a realistic environment taking into account physical phenomenon in LoRaWAN such as the capture effect and inter-spreading factor interference. The simulation results show that the proposed simulator provides a flexible and efficient environment to evaluate various network design parameters and self-management solutions as well as verify the effectiveness of distributed reinforcement-based learning algorithms for resource allocation problems in LoRaWAN.

This paper was published at the conference WCNC 2019 [15].

6.12. A Survey of Recent Extended Variants of the Traveling Salesman and Vehicle Routing Problems for Unmanned Aerial Vehicles

Participants: Ines Khoufi [Telecom SudParis], Anis Laouiti [Telecom SudParis], Cédric Adjih.

The use of Unmanned Aerial Vehicles (UAVs) is rapidly growing in popularity. Initially introduced for military purposes, over the past few years, UAVs and related technologies have successfully transitioned to a whole new range of civilian applications such as delivery, logistics, surveillance, entertainment, and so forth. They have opened new possibilities such as allowing operation in otherwise difficult or hazardous areas, for instance. For all applications, one foremost concern is the selection of the paths and trajectories of UAVs, and at the same time, UAVs control comes with many challenges, as they have limited energy, limited load capacity and are vulnerable to difficult weather conditions. Generally, efficiently operating a drone can be mathematically formalized as a path optimization problem under some constraints. This shares some commonalities with similar problems that have been extensively studied in the context of urban vehicles and it is only natural that the recent literature has extended the latter to fit aerial vehicle constraints. The knowledge of such problems, their formulation, the resolution methods proposed—through the variants induced specifically by UAVs features—are of interest for practitioners for any UAV application. Hence, in this study, we propose a review of existing literature devoted to such UAV path optimization problems, focusing specifically on the sub-class of problems that consider the mobility on a macroscopic scale. These are related to the two existing general classic ones—the Traveling Salesman Problem and the Vehicle Routing Problem. We analyze the recent literature that adapted the problems to the UAV context, provide an extensive classification and taxonomy of their problems and their formulation and also give a synthetic overview of the resolution techniques, performance metrics and obtained numerical results.

This paper was published in the journal "Drones" 2019, 3(3), 66 [5].

6.13. LoRa-MAB: Toward an Intelligent Resources Allocation Approach for LoRaWAN Networks

Participants: Duc-Tuyen Ta [LRI and Inria], Kinda Khawam [UVSQ], Samer Lahoud [ESIB], Cédric Adjih, Steven Martin [LRI, Université Paris-Saclay].

For a seamless deployment of the Internet of Things (IoT), self-managing solutions are needed to overcome the challenges of IoT, including massively dense networks and careful management of constrained resources in terms of calculation, memory, and battery. Leveraging on artificial intelligence will enable IoT devices to operate autonomously by using inherently distributed learning techniques. Fully distributed resource management will free devices from draining their limited energy by constantly communicating with a centralized controller. The present work is devoted to a specific IoT context, that of LoRaWAN, where devices communicate with the access network via ALOHA-type access and spread spectrum technology. Concurrent transmissions on different spreading factors increase the network capacity. However, the bottleneck is inevitable with the expected massive deployment of LoRa devices. To address this issue, we resort to the popular EXP3 (Exponential Weights for Exploration and Exploitation) algorithm to steer autonomously the decision of LoRa devices towards the least solicited spreading factors. Furthermore, the spreading factor selection is cast as a proportional fair optimization problem used as a benchmark for the learning-based algorithm. Extensive simulations were run in a realistic environment taking into account physical phenomena in LoRaWAN such as the capture effect and inter-spreading factor collision, as well as non-uniform device distribution. In such a realistic setting, we evaluate the performances of the EXP3.S algorithm, an efficient variant of the EXP3 algorithm, and show its relevance against the fair centralized solution and basic heuristics.

This paper was published at the conference GLOBECOM 2019 [16].

6.14. An IoT-Blockchain Architecture Based on Hyperledger Framework for Healthcare Monitoring Application

Participants: Oumaima Attia, Ines Khoufi [Telecom SudParis], Anis Laouiti [Telecom SudParis], Cédric Adjih.

Blockchains are one of the most promising technologies in the domain of the Internet of Things (IoT). At the same time, healthcare monitoring is one of IoT applications where many devices are connected, and collect data that need to be stored in a highly secure way. In this context, we focus on IoT Blockchain architectures for healthcare monitoring applications. We start our study by exploring both IoT and blockchain technologies and identify how Fabric Hyperledger is a blockchain framework that fits our application needs. In this paper, we propose a security architecture based on this framework. We validate our approach first at a design level through concrete examples, then by showing some implemented functionalities.

This paper was published at the conference NTMS 2019 [10].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. GranData:

Participants: Douglas Do Couto Teixiera, Licia Amichi, Lucas Santos de Oliveira [EMBRACE], Aline Carneiro Viana.

Since June 2014, we have a collaboration with GranData (http://grandata.com/), Buenos Aires, Argentina on traffic vs mobility modeling of smartphone users. GranData is a small company that integrates first-party and telco partner data to understand key market trends, to predict customer behavior, and to deliver business results. For the time being, the collaboration with Grandata has generated knowledge transfer. From both directions, (1) from myself to GranData, I have being transferring my knowledge in modeling and analysing human behavior in terms of mobility, encounters, and content demand, (2) from them to myself, they have advising me on issues related to machine learning and statistical methods to be used. It describes **an industrial partner's collaboration having the outcomes of our works impacting their products** (e.g., GranData data mining algorithms can be improved based on the better understanding on mobility and content consumption of mobile users) **or research/business decisions** (e.g., proved strong correlations between mobility and data traffic consumption can open new perspectives of services to telecom operators, i.e., clients of GranData).

Part of the thesis of Guangshuo Chen (endend April 2018) and of Eduardo Mucelli (ended in 2015) on data traffic analysis used telco traces provided by GranData.

7.2. Bilateral Grants with Industry

7.2.1. Nokia (ADR):

Participants: Cedric Adjih, Iman Hmedoush.

Through the common Inria-Nokia laboratory, the team is involved in the action "Network Information Theory" (ADR, "Action De Recherche"). In collaboration with Nokia, and Inria EPI MARACAS, and EPI EVA, we are working on the subject of optimization and evaluating communications for IoT networks. This includes 5G and beyond, medium-access level/random access techniques protocols and applying machine learning techniques to wireless communications.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Digicosme - Thesis - ECOMICENE

Participants: Cedric Adjih, Hirah Malik, Michel Kieffer [L2S, CNRS–CentraleSupelec–Univ Paris-Sud, Univ Paris-Saclay], Claudio Weidmann [ETIS / ENSEA - Université de Cergy-Pontoise, CNRS (UMR 8051)].

Partners: Centrale-Supelec L2S, ETIS-ENSEA

Subject : Efficient COding of Meta-information in Information-CEntric NEtworks.

8.1.2. Digicosme - Post doc - ICN-based-Vehicles

Participants: Cedric Adjih, Ines Khoufi [Telecom SudParis], Anis Laouiti [Telecom SudParis].

Partners: SAMOVAR, Telecom Sud-Paris (IPP)

Subject: In this work, the project is to design and propose a new architecture model that combines several new emerging research fields which are FANETs (Flying Ad-hoc NETworks). We will modelled a FANET problem of information gathering and distribution, reviewed related literature in [5]. We are now focusing on some mobility patterns for the FANETs in order to optimize the movement of the flying vehicles while they are enhancing the radio coverage for the VANETs and trying to improve data exchange experience between different damaged locations, using genetic algorithms. (link)

8.1.3. Digicosme - Engineer - LoRaWAN simulator

Participants: Cedric Adjih, Kinda Khawam [UVSQ], Samer Lahoud [ESIB], Steven Martin [LRI, Université Paris-Saclay].

Subject: LoRa-MAB: A Flexible Simulator for Decentralized Learning Resource Allocation in IoT Networks

The simulator is available at https://github.com/tuyenta/IoT-MAB

8.2. National Initiatives

8.2.1. Equipex FIT:

Participants: Cedric Adjih, Alexandre Abadie [Inria, SED], Emmanuel Baccelli.

Partners: Sorbonne Université, Inria (Lille, Sophia-Antipolis, Grenoble), INSA, Institut Telecom Paris, Institut Télécom Evry, LSIIT Strasbourg.

FIT (Future Internet of Things) aims to develop an experimental facility, a federated and competitive infrastructure with international visibility and a broad panel of customers. It provides this facility with a set of complementary components that enable experimentation on innovative services for academic and industrial users. The project gives french internet stakeholders a means to experiment on mobile wireless communications at the network and application layers thereby accelerating the design of advanced networking technologies for the future internet. FIT was one of 52 winning projects from the first wave of the French Ministry of Higher Education and Research's "Équipements d'Excellence" (Equipex) research grant program, in 2011.

One component of the FIT platform is the sets of IoT-LAB testbeds (see the IoT-LAB web site). These were motivated by the observation that the world is moving towards an "Internet of Things", in which most communication over networks will be between objects rather than people.

8.2.2. ANR

8.2.2.1. MITIK

Participants: Aline Carneiro Viana, Catuscia Palamidessi.

Funding instrument/scientific committee: PRC/CE25

Project acronym: MITIK

Project title: Mobility and contact traces from non-intrusive passive measurements

Duration: 2020–2023

Coordinator: Aline Carneiro Viana

Other partners: COMETE/Inria, Universite de la Rochelle, Sorbonne Universite.

Abstract: The MITIK project is a 42-month ANR project that will start in February 2020. Mitik's primary objective is the design of an entirely new methodology to help the community obtain real wireless contact traces that are non-intrusive, representative, and independent of third parties. The secondary outcome of the project is be the public release of (1) the measurement tool designed for the easy contact gathering task; (2) contact traces which are clean, processed, and privacy-preserving, i.e., protecting both the anonymity and the location privacy of the users; and (3) their spatiotemporal statistical analysis. We expect that Miti's outcomes will support non-biased research on the modeling as well as on the leveraging of wireless contact patterns.

8.2.2.2. GORILLA

Participants: Cedric Adjih, Aline Carneiro Viana, Nadjib Achir.

Funding instrument/scientific committee: Under submission to the PRC/CE25 (Phase I)

Project acronym: GORILLA

Project title: GeO-distributed pRivacy-preserving InteLLigent orchestrAtion of data-hungry Services

Duration: 2021–2024

Coordinator: Cedric Adjih

Other partners: IRIT – Toulouse INP, LS2N – IMT Atlantiquen L2TI – University Paris 13.

Abstract: The GORILLA project has been submitted to the ANR - PCR program (PHASE I). Users of mobile applications keep calling for better user privacy while getting better user experience, and this fact has become a competitive challenge for application developers. As of today, privacy is often promoted through personal storage and is sometimes opposed to cloud solutions which are nevertheless well-established. GORILLAS proposes to revisit this dilemma with the recent emergence of edge computing. The idea is to leverage edge computing as a middle ground that will act as a trusted third party that ensures privacy and confidentiality requirements. GORILLAS will design a framework that captures the user's privacy requirements, the services requirements as well as current and future users, networks, edge, and cloud operational contexts to perform privacy-persistent and QoE-aware data placement in addition to a tailored QoE-aware service computing orchestration over edge and cloud resources

8.3. European Initiatives

8.3.1. H2020 SPARTA project

Participants: Emmanuel Baccelli, Francois-Xavier Molina.

Program: H2020 SU-ICT-03-2018: Establishing and operating a pilot for a Cybersecurity Competence Network to develop and implement a common Cybersecurity Research & Innovation Roadmap

Project acronym: SPARTA

Project title: Strategic Programs for Advanced Research and Technology in Europe

Duration: 2019-2022

Participant from TRiBE: Emmanuel Baccelli, Francois-Xavier Molina

Other partners include CEA, TU Muenchen, IMT among many others

Abstract: The Sparta project is a 3-year H2020 project started in February 2019, which will put in motion a competence network on cybersecurity, with a view to shape a future EU-wide cybersecurity agency. In more details: TRiBE participates on topics around low-power IoT security, whereby RIOT is used as the base platform on top of which advances will be experimented with and made available in practice.

8.4. International Initiatives

8.4.1. Inria Project Lab RIOT-fp

Project lead: Emmanuel Baccelli

Full name: Reconcile IoT & Future-Proof Security

Partners: teams EVA, PROSECCO (Inria Paris), teams GRACE, TRiBE (Inria Saclay), team TEA, CELTIQUE (Inria Rennes), Freie Universitaet Berlin

Project Start: April 2019

Project Length: 4 years

Website: https://future-proof-iot.github.io/

Summary:

Today's Internet of Things (IoT) does not provide an acceptable tradeoff of functionality vs. risk for end-users. To improve this tradeoff, we must simultaneously

(i) enrich IoT functionality and (ii) improve IoT cyber-security with respect to diverse attack vectors. Concerning the former, RIOT is emerging as one of the major open-source software platforms for low-end IoT devices. Concerning the latter, research challenges must be addressed in various domains including secure network protocol stacks, cryptography, software execution guarantees, embedded system design. RIOT-fp is a research project on IoT cyber-security. Taking a global and

practical approach, RIOT-fp gathers partners planning a scientific agenda aiming to enhance RIOT with an array of security mechanisms. The main scientific challenges tackled by RIOT-fp are: (1) developing high-speed, high-security, low-memory IoT crypto primitives, (2) providing guarantees for software execution on low-end IoT devices, and (3) enabling secure IoT software updates and supply-chain, over the network.

8.4.2. Inria Associate Teams Not Involved in an Inria International Labs

8.4.2.1. EMBRACE

Title: Leveraging Human Behavior and Uncertainty in 5G Networks to Build Robust Resource Allocation and Services Orchestration Models

International Partners (Institution - Laboratory - Researcher):

UTFPR (Brazil) - Departamento Acadêmico de Informática (DAINF) Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEI) - Anelise Munaretto

UFG (Brazil) - Institute of Computational Mathematics and Scientific / Engineering Computing - Kleber Vieira Cardoso

UFMG (Brazil) - Dpt of Statistics - Antonio A. F. Loureiro

Start year: 2017 – Ending year: 2019

See also: https://team.inria.fr/embrace/

Abstract: EMBRACE propose une architecture novatrice pour gérer des ressources et des services opérationnels hétérogènes. EMBRACE se concentre sur les défis scientifiques liés des ensembles de données collectées dans le monde réel et décrivant le comportement du réseau des utilisateurs. En particulier, EMBRACE exploite la modélisation du comportement humain en termes de mobilité, de demande de contenu, d'intérêts communs et des interactions entre-utisateurs. En construisant des modèles d'allocation les ressources tenant compte de l'utilisateur, EMBRACE a pour objectif de diminuer l'incertitude et mieux cerner les profils humains dans les réseaux 5G. La communication D2D sera également utilisée comme service opérationnel pour gérer la croissance du trafic mobile en libérant des ressources des réseaux cellulaires, sans augmenter les coûts. La nouveauté de l'architecture réside dans les algorithmes conçus qui exploiteront les caractérisations tirés de l'analyse du comportement des utilisateurs, l'hétérogénéité du réseau, et de l'incertitude. L'évaluation par simulation et l'émulation sera également l'un des thèmes clés. Enfin, les équipes concernées (Inria Infine, UFMG, UFG, UTFPR) ont un long historique de coopération sur ces thèmes.

Nest steps: A new proposal extending the EMBRACE project was submitted in Nov. 2019. Besides, partners keep going their collaborations with two students currently visiting the team (Lucas Santos from UFMG and Felipe Fonseca from UFG) and with two researchers from UFG starting their sabbatical year from February 2020.

8.4.3. Inria International Partners

8.4.3.1. Declared Inria International Partners

1. Renewed IOTPUSH collaboration with Freie Universitaet Berlin around the long-term stay of Emmanuel Baccelli in Berlin, on research topics about the Internet of Things, RIOT and Information-Centric Networking.

8.4.3.2. Informal International Partners

1. Although the project has finished, the team keep going their collaboration with UFMG and UFG institutions, previous partners of EMBRACE project, on human behavior leveraging in 5G networks.
2. Collaboration with Mark Crovella from Boston University, where Licia Amichi will spend 5 months in an internship from March 2020. She will work on our current collaboration on the modelling and analysis of novelty-seeking preferences in human mobility.
3. Collaboration with Javier Bustos from NIC Lab/University of Chile, involving the PhD co-advising of Diego Madriaga, who is doing a joint PhD between Univ. of Chile and IPP and is working on short-term time series analysis and prediction for anticipatory Nnetworking.
4. Collaboration with Ana Aguiar from University of Porto, involving the PhD co-advising of Emanuel Lima, who is working on data offloading via mobile crowdsensing.
5. Collaboaration with Marco Fiore from IMDEA on adaptive sampling of human mobility. This collaboration involves the participation of Diego Madriaga.
6. Informal collaborations with ENSI Tunis and ENIso.

8.4.4. Participation in Other International Programs

8.4.4.1. STIC AmSud MOTIf 2017

Participant: Aline Carneiro Viana.

Program: STIC AmSud

Project title: Mobile phone sensing of human dynamics in techno-social environment

Duration: 2017-2019

Coordinators: Marton Karsai (ENS/Inria) and Jussara M. Almeida (UFMG) and Alejo Salles (Univ. of Buenos Aires)

Abstract: Information and Communication Technology (ICT) is becoming increasingly social, as demonstrated by the multitude of emerging technologies and technology platforms that facilitate social interactions, taking place as communication via telephone, text message, email, online social networks etc. At the same time, our social activities are increasingly embedded in the ICT environments that enable and enhance our ability to transact, share experiences, and maintain social relationships. One of the best ways to explore these developments is through the mining and analysis of data, which are collected through mobile phones and allow us to investigate how individuals act when embedded in a technology-enabled environment. The MOTIf project builds on the analysis and modeling of geo-localized temporally detailed but fully anonymised mobile phone call networks. These datasets allow us to address the two scientific objectives about spatiotemporal patterns of service usage of anonymised individuals to learn when, where, and what people are doing; and about the fine-grained sociodemographic structure of society and its effect on the the individual social behaviour. In other words our goal in general is to understand how individuals behave in a dynamic techno-social environment.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Prof. Kleber Vieira Cardoso and Sand Luz Correa forom UFG, Brazil, will do their sabbatical year at the TRiBE team, under Brazilian funding and in the context of the EMBRACe project. They will work with Aline C. Viana and Felipe Fonseca on trajectory reconstruction of tourists and their 5G resource optimization.

8.5.2. Internships

Felipe Fonseca is doing an internship of 3 months in our team (Nov 201-Jan 2020). He work with Aline C. Viana, Kleber V. Cardoso and Sand L. Correa on trajectory reconstruction of tourists.

Lucas Santos is doing an internship of 3 months in our team (Nov 201-Jan 2020) in the context of EMBRACE associated team. He work with Aline C. Viana and Pedro Olmo on the investigation of causalities in habits of human visits.

Douglas Teixira did an internship of 10 months our team (May 2019-Jan 2020) in the context of EMBRACE associated team. He is in cotutelle between IPP and UFMG and is co-advised by Aline C. Viana and Jussara Almeida on the limits of a context-aware predictability of human mobility.

Amina Ben Hassine did an intership of 6 months (2019) in collaboration with Ichrak Amdouni (ENSISo) and Anis Laouiti (Telecom SudParis) on the subject of "Unmanned Aerial Vehicles Path Planning Using Machine Learning" using reinforcement learning.

8.5.3. Visits to International Teams

8.5.3.1. Research Stays Abroad

Aside of working for Inria, **Emmanuel Baccelli** is also Professor at Freie Universitaet (FU) Berlin, within the context of a chair resulting of a partnership between Inria, FU Berlin and Einstein Center for Digital Future (ECDF: umbrella organization for Berlin’s technical universities). The topic of this chair is *Open and Secure IoT Ecosystem*. In this context, Emmanuel Baccelli stays at FU Berlin. See online: https://www.digital-future.berlin/en/about-us/professors/prof-dr-emmanuel-baccelli/

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

- Aline C. Viana is **Publicity co-chair** of ESWN 2020; was **Student Travel Grant co-chair** of IEEE Infocom 2019;
- **Emmanuel Baccelli** was co-chair of the RIOT Summit'19.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- Aline C. Viana is **TPC co-chair** of Shadow Algotel/Cores 2020; This will be the first time a shadow TPC will be organized for these two French conferences. A shadow TPC aims at providing an educational experience for young PhD graduates, post docs, and junior researchers by simulating a TPC meeting entitled to discuss some papers submitted to Algotel and Cores 2020.

9.1.2.2. Member of the Conference Program Committees

- Aline C. Viana is TPC member of Algotal 2020;

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Aline C. Viana is an Editorial Board member of Ad Hoc Networks, Elsevier journal since Nov 2029; Urban Computing Spring book series (http://www.springer.com/series/15552), since Feb. 2018; an Associate Editor of ACM SIGCOMM Computer Communication Review (ACM CCR), since May 2014; an Editorial Board member of Wireless Communications and Mobile Computing Open Access Journal of John Wiley&Sons and Hindawi since 2016.

9.1.3.2. Reviewer - Reviewing Activities

- **Aline C. Viana** reviewed papers for ACM SIGCOMM CCR Journal, IEEE Transaction on Mobile Computing journal, Elsevier Pervasive and Mobile Computing Journal, and Computer Communication Elsevier Journal.
- **Emmanuel Baccelli** reviewed papers for IEEE IoT Journal, IETF Routing Directory, IFIP/IEEE PEMWN.
- **Cedric Adjih** was a reviewer for ACM SIGCOMM CCR Journal.

9.1.4. Invited Talks

- **Aline C. Viana** was invited to give a seminar at (1) the GDR RSD and ASF Winter School, Pleynet, France, in Feb. 2019; (2) the AGRANDA symposium in Salta, Argentina in Sep. 2019; (3) at IFIP TMA 2019 Expert Summit, https://tma.ifip.org/2019/tma-experts/.
- **Emmanuel Baccelli** was invited to give talks at the French Ministry of Defence (Innovation Defence Lab), at the Berlin Innovation Agency Smart Cities Meetup, and at Eclipse IoT Day.

9.1.5. Standardization

1. **Emmanuel Baccelli and Cedric Adjih** have participated at several working groups at IETF during 2019.
2. **Cedric Adjih** has participated at all IETF Hackathons in 2019 (IETF 104, IETF 105, IETF 106), along with **Oumaima Attia**; they were major contributors to
   – LPWAN Hackathon on the SCHC protocol (IPv6 compression for IoT networks), see: https://github.com/openschc/openschc

  – NWCRG Hackathon on the SWIF-codec, a sliding-window forward-error correction codec, see: https://github.com/irtf-nwcrg/swif-codec

3. **Emmanuel Baccelli** has participated at several IETF hackathons during 2019.

9.1.6. Scientific Expertise

- **Aline C. Viana** served as (1) Remote evaluator of the *ERC's Starting Grant 2019*; (2) as member of *4 PhD defense committees* (2 as reviewer and 2 as examiner); (3) as examiner in *4 PhD mid-term committees* in 2019.
- **Emmanuel Baccelli** served as reviewer for 1 PhD defense committee in 2019.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- **Aline C. Viana** gave talks/seminars at conferences and workshops, on subjects related to human behavior analysis: (1) the GDR RSD and ASF Winter School, Pleynet, France, in Feb. 2019 (https://sites.google.com/site/rsdwinterschool/last-editions/asf-winter-school-2019); (2) the AGRANDA symposium in Salta, Argentina in Sep. 2019 (http://48jaiio.sadio.org.ar/simposios/AGRANDA); (3) at IFIP TMA 2019 Expert Summit, in Jun 2019 (https://tma.ifip.org/2019/tma-experts/).
- Master : **Emmanuel Baccelli**, "IoT & Security", summer semester seminar, Freie Universitaet Berlin.
- Master : **Emmanuel Baccelli**, "IoT & Security", winter semester seminar, Freie Universitaet Berlin.
- Master : **Emmanuel Baccelli**, "Operating Systems for Small Connected Devices in the Internet of Things", cours magistral, Formation PESTO Corps des Mines, Telecom ParisTech, Paris France
- Engineering School: Cédric Adjih, "Internet of Thing", 3h practical courses, Telecom SudParis
- Engineering School: Cédric Adjih, "Internet of Thing", 9h practical courses, ENSEA

9.2.2. Supervision

- PhD in progress: Licia Amichi, "Identifying and profiling novelty-seeking behavior in human mobilityModelling exploration factor of human beings", since Oct. 2018. Advisor: Aline C. Viana
- PhD in progress: Lucas Santos, "Investigating causalities in habits of human visits", since May 2018. Advisor: Aline C. Viana and Pedro Olmo
- PhD in progress: Douglas Teixeira, "Deciphering Predictability Limits in Human Mobility", since April 2018. Advisor: Aline C. Viana and Jussara Almeida.
- PhD in progress: Rafael Costa, "Human-enhanced forwarding strategies for Device-to-Device (D2D) communication", since May 2017. Advisor: Aline C. Viana and Leobino Sampaio.
- PhD in progress: Diego Madriaga, "Short-term Time Series Analysis and Prediction for Anticipatory Networking", since Jan 2019. Advisor: Aline C. Viana and Javier Bustos.
- PhD in progress: Anne Josiane Kouam Djuigne, "Detection of bypass frauds in cellular network datasets", since Nov 2019. Advisor: Aline C. Viana and Alain Tchana.
- PhD in progress: Hirah Malik, "Efficient COding of Meta-information in Information-CEntric NEtworks", since Oct. 2017. Advisors: C. Adjih, Michel Kieffer, and Claudio Weidmann
- PhD in progress: Iman Hmedoush, "Connection protocols for the 5G IoT", since Oct. 2018. Advisors: C. Adjih and Paul Mühlethaler.

9.2.3. Juries

- **Reviewer for PhD thesis committee:**Aline C. Viana was reviewer for the following PhDs: R. Teles, Industrial IoT (Univ. de Strasbourg, fin 2019). Nadjib Achir was was reviewer for the following PhDs: M. Rautu, Déploiement temporaire d'une infrastructure de communication à base de drones (Univ. de Toulouse, octobre 2019).

- **Examiner for PhD thesis committee:**C. Adjih was examiner for the committees of the following PhDs: Safan ALWAN (UPEC, LSSI, Dev. 2019); Aline C. Viana was examiner for the committees of the following PhDs: H. Mazouzi, Algorithms for Tasks Offloading on Multiple Mobile Edge Servers (Univ. Paris 13, Nov 2019); J. Loudet, Distributed and Privacy-Preserving Personal Queries on Personal Clouds (Univ. de Versailles/PETRUS Inria, Oct 2019); J. Munoz, km-scale Industrial Networking (UPMC/EVA Inria, Mar. 2019); A. Boubrima, Deployment and Scheduling of Wireless Sensor Networks for Air Pollution Monitoring (INSA-Lyon/AGORA Inria, Mar. 2019).
- **Examiner for PhD mid-term committee:**Aline C. Viana was examiner for the committees in the following mid-term juries: J. Kamal, Détection d'anomalies comportementales pour les systèmes de transport intelligents et coopératifs, (TPT, Apr. 2019); Y. Du, In-Network Collaborative Mobile Crowd Sensing: a Context-Aware Sensing Group Framework, (UPMC/MIMOVE Inria, Mar. 2019).
- **Examiner for M.Sc. thesis committee:**Aline C. Viana was examiner for the committees in the following M.Sc. jury: W. Z. Xavier, (PUC-MG, Minas Gerais, March 2019).

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

1. **Aline C. Viana** was the President of the Scientific Commission at Inria Saclay, responsible for the selection of candidates for the CORDI-S, Post-Doc and Delegation campaigns.
2. **Aline C. Viana** is the coordinator of the ANR MITIK project to start Feb 2020 and was the international coordinator of the EMBRACE associated team of Inria.
3. Together with the members of the team, **Aline C. Viana** has submitted the short proposal of the new team as well as has presented it at the BCEP and CEP, for validation. TRiBE was officially created in Jun 2019.
4. **Aline C. Viana** is a member of BCEP, evaluating Inria teams in process of creation, discussing main issues related to different scientific commissions, discussing changes in the institution.
5. **Aline C. Viana** is co-Coordinator of the mentoring program for researchers at Inria Saclay (https://project.inria.fr/mentoratscl/). The program goal is the coach of junior researchers by more experienced ones in order to provide them with a complementary perspective for their career, independently of any hierarchical link.

10. Bibliography

Major publications by the team in recent years

[1] CISCO.*Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper*, 2018

[2] FORBES.*2017 Roundup Of Internet Of Things Forecasts*, 2017

[3] GARTNER.*Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016*, 2017

Publications of the year

Articles in International Peer-Reviewed Journal

[4] G. CHEN, A. CARNEIRO VIANA, M. FIORE, C. SARRAUTE.*Complete Trajectory Reconstruction from Sparse Mobile Phone Data*, in "EPJ Data Science", October 2019 [*DOI :* 10.1140/EPJDS/S13688-019-0206-8], https://hal.inria.fr/hal-02286080

[5] I. KHOUFI, A. LAOUITI, C. ADJIH. *A survey of recent extended variants of the traveling salesman and vehicle routing problems for Unmanned Aerial Vehicles*, in "Drones", September 2019, vol. 3, n° 3, p. 66:1-30 [*DOI :* 10.3390/DRONES3030066], https://hal.inria.fr/hal-02431644

[6] T. H. SILVA, A. CARNEIRO VIANA, F. BENEVENUTO, L. VILLAS, J. SALLES, A. A. F. LOUREIRO, D. QUERCIA. *Urban Computing Leveraging Location-Based Social Network Data: a Survey*, in "ACM Computing Surveys", March 2019 [*DOI :* 10.1145/3301284], https://hal.inria.fr/hal-01972129

[7] H. TSCHOFENIG, E. BACCELLI. *Cyberphysical Security for the Masses: A Survey of the Internet Protocol Suite for Internet of Things Security*, in "IEEE Security and Privacy Magazine", September 2019, vol. 17, n° 5, p. 47-57 [*DOI :* 10.1109/MSEC.2019.2923973], https://hal.inria.fr/hal-02351892

[8] K. ZANDBERG, K. SCHLEISER, F. ACOSTA, H. TSCHOFENIG, E. BACCELLI. *Secure Firmware Updates for Constrained IoT Devices Using Open Standards: A Reality Check*, in "IEEE Access", 2019, vol. 7, p. 71907-71920 [*DOI :* 10.1109/ACCESS.2019.2919760], https://hal.inria.fr/hal-02351794

International Conferences with Proceedings

[9] L. AMICHI, A. CARNEIRO VIANA, M. CROVELLA, A. A. F. LOUREIRO. *Mobility profiling: Identifying scouters in the crowd*, in "CoNEXT '19: Proceedings of the 15th International Conference on emerging Networking EXperiments and Technologies", Orlando, United States, December 2019, p. 9-11 [*DOI :* 10.1145/3360468.3366771], https://hal.inria.fr/hal-02433795

[10] O. ATTIA, I. KHOUFI, A. LAOUITI, C. ADJIH. *An IoT-blockchain architecture based on hyperledger framework for health care monitoring application*, in "NTMS 2019: 10th IFIP International Conference on New Technologies, Mobility and Security", Canary islands, Spain, 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE Computer Society, 2019, p. 1-5 [*DOI :* 10.1109/NTMS.2019.8763849], https://hal.inria.fr/hal-02434834

[11] J. B. BORGES, H. S. RAMOS, R. A. F. MINI, A. CARNEIRO VIANA, A. A. F. LOUREIRO. *The quest for sense: Physical phenomena classification in the Internet of Things*, in "ISIoT 2019 - 1st International Workshop on Intelligent Systems for IoT", Santorini, Greece, May 2019, https://hal.inria.fr/hal-02165145

[12] A. DI LUZIO, A. C. VIANA, K. CHATZIKOKOLAKIS, G. DIKOV, C. PALAMIDESSI, J. STEFA. *Catch Me If You Can: How Geo-indistinguishability Affects Utility in Mobility-based Geographic Datasets*, in "Proceedings of the 3rd ACM SIGSPATIAL International Workshop", Chicago, United States, ACM Press, 2019, p. 1-10 [*DOI :* 10.1145/3356994.3365498], https://hal.archives-ouvertes.fr/hal-02423337

[13] F. FONSECA, L. MAMATAS, A. CARNEIRO VIANA, S. L. CORREA, K. V. CARDOSO. *Personalized Travel Itineraries with Multi-access Edge Computing Touristic Services*, in "IEEE Globecom 2019 - IEEE Global Communications Conference", Big Island, Hawaii, United States, December 2019, https://hal.inria.fr/hal-02285929

[14] L. SANTOS DE OLIVEIRA, P. O. VAZ DE MELO, A. CARNEIRO VIANA. *Measuring Power Relations Among Locations From Mobility Data*, in "ACM MobiWac 2019 - 17th ACM International Symposium on Mobility Management and Wireless Access", Miami, United States, November 2019, https://hal.inria.fr/hal-02302490

[15] D.-T. TA, K. KHAWAM, S. LAHOUD, C. ADJIH, S. MARTIN. *LoRa-MAB: A Flexible Simulator for Decentralized Learning Resource Allocation in IoT Networks*, in "WMNC 2019 - 12th

IFIP Wireless and Mobile Networking Conference", Paris, France, IEEE, September 2019, p. 55-62 [*DOI :* 10.23919/WMNC.2019.8881393], https://hal.inria.fr/hal-02431653

[16] D.-T. TA, K. KHAWAM, S. LAHOUD, C. ADJIH, S. MARTIN.*LoRa-MAB: Toward an Intelligent Resource Allocation Approach for LoRaWAN*, in "GLOBECOM 2019 - IEEE Global Communications Conference", Waikoloa, United States, December 2019, https://hal.inria.fr/hal-02431672

[17] D. D. C. TEIXEIRA, A. CARNEIRO VIANA, M. S. ALVIM, J. M. ALMEIDA.*Deciphering Predictability Limits in Human Mobility*, in "ACM SIGSPATIAL 2019 - 27th International Conference on Advances in Geographic Information Systems", Chicago, United States, November 2019 [*DOI :* 10.1145/3347146.3359093], https://hal.inria.fr/hal-02286128

Project-Team TROPICAL

Tropical methods: structures, algorithms and interactions

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization and control of dynamic systems

Table of contents

Project-Team TROPICAL

Creation of the Team: 2016 January 01, updated into Project-Team: 2018 July 01

Keywords:

Computer Science and Digital Science:

A1.2.4. - QoS, performance evaluation
A2.3.3. - Real-time systems
A2.4. - Formal method for verification, reliability, certification
A6.2.5. - Numerical Linear Algebra
A6.2.6. - Optimization
A6.4.6. - Optimal control
A7.2.4. - Mechanized Formalization of Mathematics
A8.1. - Discrete mathematics, combinatorics
A8.2.1. - Operations research
A8.2.3. - Calculus of variations
A8.3. - Geometry, Topology
A8.9. - Performance evaluation
A8.11. - Game Theory
A9.6. - Decision support

Other Research Topics and Application Domains:

B4.3. - Renewable energy production
B4.4. - Energy delivery
B4.4.1. - Smart grids
B6.6. - Embedded systems
B8.4. - Security and personal assistance
B8.4.1. - Crisis management

1. Team, Visitors, External Collaborators

Research Scientists

Stéphane Gaubert [Team leader, Inria, Senior Researcher, HDR]
Marianne Akian [Inria, Senior Researcher, HDR]
Xavier Allamigeon [Corps des Mines, under secondment, Inria, Researcher]
Yang Qi [Inria, Starting Research Position, from Nov 2019]
Cormac Walsh [Inria, Researcher]

Technical Staff

Baptiste Colin [Inria, Engineer, from May 2019]

PhD Students

Marin Boyet [Inria, PhD Student]
Maël Forcier [École Nationale des Ponts et Chaussées, PhD Student, from Sep 2019]
Maxime Grangereau [EDF Lab (Cifre)]
Paulin Jacquot [EDF Lab (Cifre)]
Omar Saadi [École polytechnique, PhD Student]
Duy Nghi Benoît Tran [Univ Paris-Est Marne La Vallée, PhD Student]

Post-Doctoral Fellow

Aurélien Sagnier [École polytechnique, Post-Doctoral Fellow, until Aug 2019]

Visiting Scientist

Sara Vannucci [PhD student, Univ. Salerno, from Mar 2019 until Jun 2019]

Administrative Assistants

Hanadi Dib [Inria, Administrative Assistant, until Apr 2019]

Ines Dumontier [Inria, Administrative Assistant, until May 2019]

2. Overall Objectives

2.1. Introduction

The project develops tropical methods motivated by applications arising in decision theory (deterministic and stochastic optimal control, game theory, optimization and operations research), in the analysis or control of classes of dynamical systems (including timed discrete event systems and positive systems), in the verification of programs and systems, and in the development of numerical algorithms. Tropical algebra tools are used in interaction with various methods, coming from convex analysis, Hamilton–Jacobi partial differential equations, metric geometry, Perron-Frobenius and nonlinear fixed-point theories, combinatorics or algorithmic complexity. The emphasis of the project is on mathematical modelling and computational aspects.

The subtitle of the *Tropical* project, namely, “structures, algorithms, and interactions”, refers to the spirit of our research, including a methodological component, computational aspects, and finally interactions with other scientific fields or real world applications, in particular through mathematical modelling.

2.2. Scientific context

Tropical algebra, geometry, and analysis have enjoyed spectacular development in recent years. Tropical structures initially arose to solve problems in performance evaluation of discrete event systems [56], combinatorial optimization [61], or automata theory [107]. They also arose in mathematical physics and asymptotic analysis [96], [93]. More recently, these structures have appeared in several areas of pure mathematics, in particular in the study of combinatorial aspects of algebraic geometry [82], [121], [110], [87], in algebraic combinatorics [75], and in arithmetics [67]. Also, further applications of tropical methods have appeared, including optimal control [100], program invariant computation [48] and timed systems verification [95], and zero-sum games [2].

The term ‘tropical’ generally refers to algebraic structures in which the laws originate from optimization processes. The prototypical tropical structure is the max-plus semifield, consisting of the real numbers, equipped with the maximum, thought of as an additive law, and the addition, thought of as a multiplicative law. Tropical objects appear as limits of classical objects along certain deformations (“log-limits sets” of Bergman, “Maslov dequantization”, or “Viro deformation”). For this reason, the introduction of tropical tools often yields new insights into old familiar problems, leading either to counterexamples or to new methods and results; see for instance [121], [102]. In some applications, like optimal control, discrete event systems, or static analysis of programs, tropical objects do not appear through a limit procedure, but more directly as a modelling or computation/analysis tool; see for instance [117], [56], [85], [62].

Tropical methods are linked to the fields of positive systems and of metric geometry [104], [12]. Indeed, tropically linear maps are monotone (a.k.a. order-preserving). They are also nonexpansive in certain natural metrics (sup-norm, Hopf oscillation, Hilbert’s projective metric, ...). In this way, tropical dynamical systems appear to be special cases of nonexpansive, positive, or monotone dynamical systems, which are studied as part of linear and non-linear Perron-Frobenius theory [94], [3]. Such dynamical systems are of fundamental importance in the study of repeated games [101]. Monotonicity properties are also essential in the understanding of the fixed points problems which determine program invariants by abstract interpretation [69]. The latter problems are actually somehow similar to the ones arising in the study of zero-sum games; see [7]. Moreover, positivity or monotonicity methods are useful in population dynamics, either in a discrete space setting [119] or in a PDE setting [57]. In such cases, solving tropical problems often leads to solutions or combinatorial insights on classical problems involving positivity conditions (e.g., finding equilibria of dynamical systems with nonnegative coordinates, understanding the qualitative and quantitative behavior of growth rates / Floquet eigenvalues [10], etc). Other applications of Perron-Frobenius theory originate from quantum information and control [109], [115].

3. Research Program

3.1. Optimal control and zero-sum games

The dynamic programming approach allows one to analyze one or two-player dynamic decision problems by means of operators, or partial differential equations (Hamilton–Jacobi or Isaacs PDEs), describing the time evolution of the value function, i.e., of the optimal reward of one player, thought of as a function of the initial state and of the horizon. We work especially with problems having long or infinite horizon, modelled by stopping problems, or ergodic problems in which one optimizes a mean payoff per time unit. The determination of optimal strategies reduces to solving nonlinear fixed point equations, which are obtained either directly from discrete models, or after a discretization of a PDE.

The geometry of solutions of optimal control and game problems Basic questions include, especially for stationary or ergodic problems, the understanding of existence and uniqueness conditions for the solutions of dynamic programming equations, for instance in terms of controllability or ergodicity properties, and more generally the understanding of the structure of the full set of solutions of stationary Hamilton–Jacobi PDEs and of the set of optimal strategies. These issues are already challenging in the one-player deterministic case, which is an application of choice of tropical methods, since the Lax-Oleinik semigroup, i.e., the evolution semigroup of the Hamilton-Jacobi PDE, is a linear operator in the tropical sense. Recent progress in the deterministic case has been made by combining dynamical systems and PDE techniques (weak KAM theory [72]), and also using metric geometry ideas (abstract boundaries can be used to represent the sets of solutions [86], [4]). The two player case is challenging, owing to the lack of compactness of the analogue of the Lax-Oleinik semigroup and to a richer geometry. The conditions of solvability of ergodic problems for games (for instance, solvability of ergodic Isaacs PDEs), and the representation of solutions are only understood in special cases, for instance in the finite state space case, through tropical geometry and non-linear Perron-Frobenius methods [38], [41], [3].

Algorithmic aspects: from combinatorial algorithms to the attenuation of the curse of dimensionality Our general goal is to push the limits of solvable models by means of fast algorithms adapted to large scale instances. Such instances arise from discrete problems, in which the state space may so large that it is only accessible through local oracles (for instance, in some web ranking applications, the number of states may be the number of web pages) [73]. They also arise from the discretization of PDEs, in which the number of states grows exponentially with the number of degrees of freedom, according to the "curse of dimensionality". A first line of research is the development of *new approximation methods for the value function*. So far, classical approximations by linear combinations have been used, as well as approximation by suprema of linear or quadratic forms, which have been introduced in the setting of dual dynamic programming and of the so called "max-plus basis methods" [74]. We believe that more concise or more accurate approximations may be obtained by unifying these methods. Also, some max-plus basis methods have been shown to *attenuate the curse of dimensionality* for very special problems (for instance involving switching) [97], [78]. This suggests that the complexity of control or games problems may be measured by more subtle quantities that the mere number of states, for instance, by some forms of metric entropy (for example, certain large scale problems have a low complexity owing to the presence of decomposition properties, "highway hierarchies", etc.). A second line of of our research is the development of *combinatorial algorithms*, to solve large scale zero-sum two-player problems with discrete state space. This is related to current open problems in algorithmic game theory. In particular, the existence of polynomial-time algorithms for games with ergodic payment is an open question. See e.g. [43] for a polynomial time average complexity result derived by tropical methods. The two lines of research are related, as the understanding of the geometry of solutions allows to develop better approximation or combinatorial algorithms.

3.2. Non-linear Perron-Frobenius theory, nonexpansive mappings and metric geometry

Several applications (including population dynamics [10] and discrete event systems [56], [64], [46]) lead to studying classes of dynamical systems with remarkable properties: preserving a cone, preserving an order, or being nonexpansive in a metric. These can be studied by techniques of non-linear Perron-Frobenius theory [3] or metric geometry [11]. Basic issues concern the existence and computation of the "escape rate" (which determines the throughput, the growth rate of the population), the characterizations of stationary regimes (non-linear fixed points), or the study of the dynamical properties (convergence to periodic orbits). Nonexpansive mappings also play a key role in the "operator approach" to zero-sum games, since the one-day operators of games are nonexpansive in several metrics, see [8].

3.3. Tropical algebra and convex geometry

The different applications mentioned in the other sections lead us to develop some basic research on tropical algebraic structures and in convex and discrete geometry, looking at objects or problems with a "piecewise-linear " structure. These include the geometry and algorithmics of tropical convex sets [49], [40], tropical semialgebraic sets [52], the study of semi-modules (analogues of vector spaces when the base field is replaced by a semi-field), the study of systems of equations linear in the tropical sense, investigating for instance the analogues of the notions of rank, the analogue of the eigenproblems [42], and more generally of systems of tropical polynomial equations. Our research also builds on, and concern, classical convex and discrete geometry methods.

3.4. Tropical methods applied to optimization, perturbation theory and matrix analysis

Tropical algebraic objects appear as a deformation of classical objects thought various asymptotic procedures. A familiar example is the rule of asymptotic calculus,

$$e^{-a/\epsilon} + e^{-b/\epsilon} \asymp e^{-\min(a,b)/\epsilon} \quad , \qquad e^{-a/\epsilon} \times e^{-b/\epsilon} = e^{-(a+b)/\epsilon} \quad , \tag{7}$$

when $\epsilon \to 0^+$. Deformations of this kind have been studied in different contexts: large deviations, zero-temperature limits, Maslov's "dequantization method" [96], non-archimedean valuations, log-limit sets and Viro's patchworking method [122], etc.

This entails a relation between classical algorithmic problems and tropical algorithmic problems, one may first solve the $\epsilon = 0$ case (non-archimedean problem), which is sometimes easier, and then use the information gotten in this way to solve the $\epsilon = 1$ (archimedean) case.

In particular, tropicalization establishes a connection between polynomial systems and piecewise affine systems that are somehow similar to the ones arising in game problems. It allows one to transfer results from the world of combinatorics to "classical" equations solving. We investigate the consequences of this correspondence on complexity and numerical issues. For instance, combinatorial problems can be solved in a robust way. Hence, situations in which the tropicalization is faithful lead to improved algorithms for classical problems. In particular, scalings for the polynomial eigenproblems based on tropical preprocessings have started to be used in matrix analysis [80], [84].

Moreover, the tropical approach has been recently applied to construct examples of linear programs in which the central path has an unexpectedly high total curvature [44], and it has also led to positive polynomial-time average case results concerning the complexity of mean payoff games. Similarly, we are studying semidefinite programming over non-archimedean fields [52], [51], with the goal to better understand complexity issues in classical semidefinite and semi-algebraic programming.

4. Application Domains

4.1. Discrete event systems (manufacturing systems, networks)

One important class of applications of max-plus algebra comes from discrete event dynamical systems [56]. In particular, modelling timed systems subject to synchronization and concurrency phenomena leads to studying dynamical systems that are non-smooth, but which have remarkable structural properties (nonexpansiveness in certain metrics , monotonicity) or combinatorial properties. Algebraic methods allow one to obtain analytical expressions for performance measures (throughput, waiting time, etc). A recent application, to emergency call centers, can be found in [46].

4.2. Optimal control and games

Optimal control and game theory have numerous well established applications fields: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In most of these applications, one needs either to derive analytical or qualitative properties of solutions, or design exact or approximation algorithms adapted to large scale problems.

4.3. Operations Research

We develop, or have developed, several aspects of operations research, including the application of stochastic control to optimal pricing, optimal measurement in networks [112]. Applications of tropical methods arise in particular from discrete optimization [62], [63], scheduling problems with and-or constraints [103], or product mix auctions [120].

4.4. Computing program and dynamical systems invariants

A number of programs and systems verification questions, in which safety considerations are involved, reduce to computing invariant subsets of dynamical systems. This approach appears in various guises in computer science, for instance in static analysis of program by abstract interpretation, along the lines of P. and R. Cousot [69], but also in control (eg, computing safety regions by solving Isaacs PDEs). These invariant sets are often sought in some tractable effective class: ellipsoids, polyhedra, parametric classes of polyhedra with a controlled complexity (the so called "templates" introduced by Sankaranarayanan, Sipma and Manna [113]), shadows of sets represented by linear matrix inequalities, disjunctive constraints represented by tropical polyhedra [48], etc. The computation of invariants boils down to solving large scale fixed point problems. The latter are of the same nature as the ones encountered in the theory of zero-sum games, and so, the techniques developed in the previous research directions (especially methods of monotonicity, nonexpansiveness, discretization of PDEs, etc) apply to the present setting, see e.g. [76], [81] for the application of policy iteration type algorithms, or for the application for fixed point problems over the space of quadratic forms [7]. The problem of computation of invariants is indeed a key issue needing the methods of several fields: convex and nonconvex programming, semidefinite programming and symbolic computation (to handle semialgebraic invariants), nonlinear fixed point theory, approximation theory, tropical methods (to handle disjunctions), and formal proof (to certify numerical invariants or inequalities).

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Notable article

The results of the article [45], providing an unexpected counter example to the "continuous analogue of the Hirsch conjecture", showing that log-barrier interior point methods are not strongly polynomial, have been discussed by Jesus De Loera in his survey of recent advances on Linear Programming, "Algebraic and Topological Tools in Linear Optimization", Notices de l'AMS (volume 66, number 7, 2019, especially pp. 1028-1032.

5.1.2. Awards

Maxime Grangereau (PhD student) has been laureate of the programme "Siebel Scholar 2020", https://twitter.com/polytechnique/status/1177111371835695104

6. New Software and Platforms

6.1. Coq-Polyhedra

KEYWORDS: Coq - Polyhedra - Automated theorem proving - Linear optimization

SCIENTIFIC DESCRIPTION: Coq-Polyhedra is a library providing a formalization of convex polyhedra in the Coq proof assistant. While still in active development, it provides an implementation of the simplex method, and already handles the basic properties of polyhedra such as emptiness, boundedness, membership. Several fundamental results in the theory of convex polyhedra, such as Farkas Lemma, duality theorem of linear programming, and Minkowski Theorem, are also formally proved.

The formalization is based on the Mathematical Components library, and makes an extensive use of the boolean reflection methodology.

FUNCTIONAL DESCRIPTION: Coq-Polyhedra is a library which aims at formalizing convex polyhedra in Coq

NEWS OF THE YEAR: Coq-Polyhedra now provides most of the basic operations on polyhedra. They are expressed on a quotient type that avoids reasoning with particular inequality representations. They include : * the construction of elementary polyhedra (half-spaces, hyperplanes, affine spaces, orthants, simplices, etc) * basic operations such as intersection, projection (thanks to the formalization of the Fourier-Motzkin algorithm), image under linear functions, computations of convex hulls, finitely generated cones, etc. * computation of affine hulls of polyhedra, as well as their dimension

Thanks to this, we have made huge progress on the formalization of the combinatorics of polyhedra. The poset of faces, as well as its fundamental properties (lattice, gradedness, atomicity and co-atomicity, etc) are now formalized. The manipulation of the faces is based on an extensive use of canonical structures, that allows to get the most appropriate inequality representations for reasoning. In this way, we arrive at very concise and elegant proofs, closer to the pen-and-paper ones.

- Participants: Xavier Allamigeon, Vasileios Charisopoulos, Ricardo Katz and Pierre-Yves Strub
- Partners: CIFASIS - Ecole Polytechnique
- Contact: Xavier Allamigeon
- Publications: A Formalization of Convex Polyhedra Based on the Simplex Method - A Formalization of Convex Polyhedra Based on the Simplex Method - First steps in the formalization of convex polyhedra in Coq
- URL: https://github.com/nhojem/Coq-Polyhedra

6.2. EmergencyEval

KEYWORDS: Dynamic Analysis - Simulation - Ocaml - Emergency - Firefighters - Police

SCIENTIFIC DESCRIPTION: This software aims at enabling the definition of a Petri network execution semantic, as well as the instanciation and execution of said network using the aforedefined semantic.

The heart of the project dwells in its kernel which operates the step-by-step execution of the network, obeying rules provided by an oracle. This user-defined and separated oracle computes the information necessary to the kernel for building the next state using the current state. The base of our software is the framework for the instanciation and execution of Petri nets, without making assumptions regarding the semantic.

In the context of the study of the dynamics of emergency call centers, a second part of this software is the definition and implementation of the semantic of call centers modelized as Petri nets, and more specifically timed prioritized Petri nets. A module interoperating with the kernel enables to include all the operational specificities of call centers (urgency level, discriminating between operators and callers ...) while guaranteeing the genericity of the kernal which embeds the Petri net formalism as such.

FUNCTIONAL DESCRIPTION: In order to enable the quantitative study of the throughput of calls managed by emergency center calls and the assesment of various organisationnal configurations considered by the stakeholders (firefighters, police, medical emergency service of the 75, 92, 93 and 94 French departments), this software modelizes their behaviours by resorting to extensions of the Petri net formalism. Given a call transfer protocol in a call center, which corresponds to a topology and an execution semantic of a Petri net, the software generates a set of entering calls in accord with the empirically observed statistic ditributions (share of very urgent calls, conversation length), then simulates its management by the operators with respect to the defined protocol. Transitional regimes phenomenons (peak load, support) which are not yet handled by mathematical analysis could therefore be studied. The ouput of the software is a log file which is an execution trace of the simulation featuring extensive information in order to enable the analysis of the data for providing simulation-based insights for decision makers.

The software relies on a Petri net simulation kernel designed to be as modular and adaptable as possible, fit for simulating other Petri-net related phenomenons, even if their semantic differ greatly.

- Participants: Baptiste Colin and Xavier Allamigeon
- Contact: Baptiste Colin

7. New Results

7.1. Optimal control and zero-sum games

7.1.1. Fixed points of order preserving homogeneous maps and zero-sum games

Participants: Marianne Akian, Stéphane Gaubert.

In a series of joint works with Antoine Hochart, applied methods of non-linear fixed point theory to zero-sum games.

A key issue is the solvability of the ergodic equation associated to a zero-sum game with finite state space, i.e., given a dynamic programming operator T associated to an undiscounted problem, one looks for a vector u, called the bias, and for a scalar λ, the ergodic constant, such that $T(u) = \lambda e + u$. The bias vector is of interest as it allows to determine optimal stationnary strategies.

In [41], we studied zero-sum games with perfect information and finite action spaces, and showed that the set of payments for which the bias vector is not unique (up to an additive constant) coincides with the union of lower dimensional cells of a polyhedral complex, in particular, the bias vector is unique, generically. We provided an application to perturbation schemes in policy iteration.

In [14], we apply game theory methods to the study of the nonlinear eigenproblem for homogeneous order preserving self maps of the interior of the cone. We show that the existence and uniqueness of an eigenvector is governed by combinatorial conditions, involving dominions (sets of states "controlled" by one of the two players). In this way, we characterize the situation in which the existence of an eigenvector holds independently of perturbations, and we solve an open problem raised in [77].

7.1.2. Nonlinear fixed point methods to compute joint spectral raddi of nonnegative matrices

Participant: Stéphane Gaubert.

In [21], we introduce a non-linear fixed point method to approximate the joint spectral radius of a finite set of nonnegative matrices. We show in particular that the joint spectral radius is the limit of the eigenvalues of a family of non-linear risk-sensitive type dynamic programming operators. We develop a projective version of Krasnoselskii-Mann iteration to solve these eigenproblems, and report experimental results on large scale instances (several matrices in dimensions of order 1000 within a minute).

7.1.3. Probabilistic and max-plus approximation of Hamilton-Jacobi-Bellman equations

Participant: Marianne Akian.

We consider fully nonlinear Hamilton-Jacobi-Bellman equations associated to diffusion control problems with finite horizon involving a finite set-valued (or switching) control and possibly a continuum-valued control. In [36], we constructed a lower complexity probabilistic numerical algorithm by combining the idempotent expansion properties obtained by McEneaney, Kaise and Han [92], [98] for solving such problems with a numerical probabilistic method such as the one proposed by Fahim, Touzi and Warin [70] for solving some fully nonlinear parabolic partial differential equations, when the volatility does not oscillate too much. In [37] and [27], we improved the method of Fahim, Touzi and Warin by introducing probabilistic schemes which are monotone without any restrictive condition, allowing one to solve fully nonlinear parabolic partial differential equations with general volatilities. We studied the convergence and obtain error estimates when the parameters and the value function are bounded.

7.1.4. Tropical-SDDP algorithms for stochastic control problems involving a switching control

Participants: Marianne Akian, Duy Nghi Benoît Tran.

The PhD thesis of Benoît Tran, supervised by Jean-Philippe Chancelier (ENPC) and Marianne Akian concerns the numerical solution of the dynamic programming equation of discrete time stochastic control problems.

Several methods have been proposed in the litterature to bypass the curse of dimensionality difficulty of such an equation, by assuming a certain structure of the problem. Examples are the max-plus based method of McEneaney [99], [100], the stochastic max-plus scheme proposed by Zheng Qu [108], the stochastic dual dynamic programming (SDDP) algorithm of Pereira and Pinto [105], the mixed integer dynamic approximation scheme of Philpott, Faisal and Bonnans [55], the probabilistic numerical method of Fahim, Touzi and Warin [70]. We propose to associate and compare these methods in order to solve more general structures.

In a first work [35], see also [24], we build a common framework for both the SDDP and a discrete time and finite horizon version of Zheng Qu's algorithm for deterministic problems involving a finite set-valued (or switching) control and a continuum-valued control. We propose an algorithm that generates monotone approximations of the value function as a pointwise supremum, or infimum, of basic (affine or quadratic for example) functions which are randomly selected. We give sufficient conditions that ensure almost sure convergence of the approximations to the value function. More recently, we study generalizations of these algorithms to the case of stochastic optimal control problems.

In a recent work, we introduce and study an entropic relaxation of the Nested Distance introduced by Pflug [106].

7.1.5. A variance reduction deflated value iteration algorithm to solve ergodic games

Participants: Marianne Akian, Stéphane Gaubert, Omar Saadi.

Recently, Sidford et al. introduced in [116] a variance reduced value iteration algorithm to solve discounted Markov decision processes. In [25], in a joint work work with Zheng Qu (Hong Kong University), we extended this algorithm to the ergodic (mean payoff) case, and also to the two-player case, exploiting techniques from non-linear spectral theory [39] and variational analysis. The deterministic version of this algorithm also yields a new method (alternative to relative value iteration) to solve ergodic problems.

7.2. Non-linear Perron-Frobenius theory, nonexpansive mappings and metric geometry

7.2.1. Order isomorphisms and antimorphisms on cones

Participant: Cormac Walsh.

We have been studying non-linear operators on ordered vector spaces that preserve or reverse the order structure. A bijective map that preserves the order in both directions is called an order isomorphism, and one that reverse the order in both directions is called an order antimorphism. These maps are closely related to the isometries of the Hilbert and Thompson metrics on the interior of the cone of positive elements.

The study of the order isomorphisms of an ordered vector space goes back to Alexandrov and Zeeman, who considered maps preserving the light cone that arises in special relativity. This work was extended to more general cones by Rothaus; Noll and Schäffer; and Artstein-Avidan and Slomka. It was shown, in the finite-dimensional case, that all isomorphisms are affine if the cone has no one-dimensional factors. There are also some results in infinite dimension—however these are unsatisfactory because of the strong assumptions that must be made in order to get the finite-dimensional techniques to work. For example, a typical assumption is that the positive cone is the convex hull of its extreme rays, which is overly restrictive in infinite dimension.

In a recent preprint [34], we broaden the scope of these results, requiring only very mild assumptions, namely that the spaces involved are *complete order unit spaces*. These are ordered vector spaces whose cone of positive elements is Archimedean, and that have an order unit, such that the norm induced by this order unit is complete. We show that the existence of an order isomorphism between two such spaces implies that they are in fact linearly isomorphic as ordered vector spaces.

In addition, we introduce a necessary and sufficient criterion for all order isomorphisms on a complete order-unit space to be affine. This criterion is in terms of the geometry of the dual cone. In the current setting, the dual cone has a cross-section called the state space, whose extreme points are called pure states. The closure of the set of pure states is known as the pure state space. The criterion is then that the union of the supports of the affine dependencies supported by the pure state space is dense in the pure state space.

7.2.2. Generalization of the Hellinger distance

Participant: Stéphane Gaubert.

In [58] (joint work with Rajendra Bhatia of Ashoka University and Tanvi Jain, Indian Statistic Institute, New Delhi), we study some generalizations of the Hellinger distance to the space of positive definite matrices.

7.2.3. Spectral inequalities for nonnegative tensors and their tropical analogues

Participant: Stéphane Gaubert.

In [30] (joint work with Shmuel Friedland, University of Illinois at Chicago) we extend some characterizations and inequalities for the eigenvalues of nonnegative matrices, such as Donsker-Varadhan, Friedland-Karlin, Karlin-Ost inequalities, to nonnegative tensors. These inequalities are related to a correspondence between nonnegative tensors and ergodic control: the logarithm of the spectral radius of a tensor is given by the value of an ergodic problem in which instantaneous payments are given by a relative entropy. Some of these inequalities involve the tropical spectral radius, a limit of the spectral radius which we characterize combinatorially as the value of an ergodic Markov decision process.

7.3. Tropical algebra and convex geometry

7.3.1. Formalizing convex polyhedra in Coq

Participant: Xavier Allamigeon.

This work is joint with Ricardo Katz (Conicet, Argentina) and Pierre-Yves Strub (LIX, Ecole Polytechnique).

In [54], we have made the first steps of a formalization of the theory of convex polyhedra in the proof assistant Coq. The originality of our approach lies in the fact that our formalization is carried out in an effective way, in the sense that the basic predicates over polyhedra (emptiness, boundedness, membership, etc) are defined by means of Coq programs. All these predicates are then proven to correspond to the usual logical statements. The latter take the form of the existence of certificates: for instance, the emptiness of a polyhedron is shown to be equivalent to the existence of a certificate *a la* Farkas. This equivalence between Boolean predicates and formulas living in the kind `Prop` is implemented by using the boolean reflection methodology, and the supporting tools provided by the Mathematical Components library and its tactic language. The benefit of the effective nature of our approach is demonstrated by the fact that we easily arrive at the proof of important results on polyhedra, such as several versions of Farkas Lemma, duality theorem of linear programming, separation from convex hulls, Minkowski Theorem, etc.

Our effective approach is made possible by implementing the simplex method inside Coq, and proving its correctness and termination. Two difficulties need to be overcome to formalize it. On the one hand, we need to deal with its termination. More precisely, the simplex method iterates over the so-called bases. Its termination depends on the specification of a pivoting rule, whose aim is to determine, at each iteration, the next basis. In this work, we have focused on proving that the lexicographic rule ensures termination. On the other hand, the simplex method is actually composed of two parts. The part that we previously described, called Phase II, requires an initial basis to start with. Finding such a basis is the purpose of Phase I. It consists in building an extended problem (having a trivial initial basis), and applying to it Phase II. Both phases need to be formalized to obtain a fully functional algorithm.

The most recent advances on the project are described in the software section.

7.3.2. Tropical totally positive matrices and planar networks

Participant: Stéphane Gaubert.

In [79] (joint work with Adi Niv) we characterized the tropical analogues of totally positive and totally non-negative matrices, i.e, the images by the valuation of the corresponding classes of matrices over a non-archimedean field. We showed in particular that tropical totally positive matrices essentially coincide with the Monge matrices (defined by the positivity of 2×2 tropical minors), arising in optimal transport, and compare the set of tropical totally positive matrices with the tropicalization of the totally positive Grassmannian. A fundamental property of classical totally positive matrices is their representation as weight matrices of planar network; in the recent work [31], we studied the tropical analogue of this property.

7.3.3. Linear algebra over systems

Participants: Marianne Akian, Stéphane Gaubert.

In a joint work with Louis Rowen (Univ. Bar Ilan), we study linear algebra and convexity properties over "systems". The latter provide a general setting encompassing extensions of the tropical semifields and hyperfields. A first account of this work was presented by Marianne Akian and Louis Rowen at the SIAM conference on applied algebraic geometry, in Bern.

7.3.4. Ambitropical convexity and Shapley retracts

Participants: Marianne Akian, Stéphane Gaubert.

Closed tropical convex cones are the most basic examples of modules over the tropical semifield. They coincide with sub-fixed-point sets of Shapley operators – dynamic programming operators of zero-sum games. We study a larger class of cones, which we call "ambitropical" as it includes both tropical cones and their duals. Ambitropical cones can be defined as lattices in the order induced by Rn. Closed ambitropical cones are precisely the fixedpoint sets of Shapley operators. They are characterized by a property of best co-approximation arising from the theory of nonexpansive retracts of normed spaces. Finitely generated ambitropical cones arise when considering Shapley operators of deterministic games with finite action spaces. Finitely generated ambitropical cones are special polyhedral complexes whose cells are alcoved poyhedra, and locally, they are in bijection with order preserving retracts of the Boolean cube This is a joint work with Sara Vannucci (invited PhD student from Salerno university). A first account of this work was presented by Stéphane Gaubert at the JAMI Workshop, Riemann-Roch theorem in characteristic one and related topics, in Baltimore.

7.3.5. Volume and integer points of tropical polytopes

Participant: Stéphane Gaubert.

We investigate in [20] (joint work with Marie McCaig) the volume of tropical polytopes, as well as the number of integer points contained in integer polytopes. We proved that even approximating these values for a tropical polytope given by its vertices is hard, with no approximation algorithm with factor $2^{\mathrm{poly}(m,n)}$ existing unless $\mathrm{P} = \mathrm{NP}$.

7.4. Tropical methods applied to optimization, perturbation theory and matrix analysis

7.4.1. Tropicalization of semidefinite programming and its relation with stochastic games

Participants: Xavier Allamigeon, Stéphane Gaubert.

Semidefinite programming consists in optimizing a linear function over a spectrahedron. The latter is a subset of $\mathbb{R}^n$ defined by linear matrix inequalities, i.e., a set of the form

$$\left\{x \in \mathbb{R}^n : Q^{(0)} + x_1 Q^{(1)} + \cdots + x_n Q^{(n)} \succeq 0\right\}$$

where the $Q^{(k)}$ are symmetric matrices of order m, and $\succeq$ denotes the Loewner order on the space of symmetric matrices. By definition, $X \succeq Y$ if and only if $X - Y$ is positive semidefinite.

Semidefinite programming is a fundamental tool in convex optimization. It is used to solve various applications from engineering sciences, and also to obtain approximate solutions or bounds for hard problems arising in combinatorial optimization and semialgebraic optimization.

A general issue in computational optimization is to develop combinatorial algorithms for semidefinite programming. Indeed, semidefinite programs are usually solved via interior point methods. However, the latter provide an approximate solution in a polynomial number of iterations, provided that a strictly feasible initial solution. Semidefinite programming becomes a much harder matter if one requires an exact solution. The feasibility problem belongs to $\mathsf{NP}_{\mathbb{R}} \cap \mathsf{coNP}_{\mathbb{R}}$, where the subscript $\mathbb{R}$ refers to the BSS model of computation. It is not known to be in NP in the bit model.

The PhD thesis of Mateusz Skomra [118] dealt about semidefinite programming, in the case where the field $\mathbb{R}$ is replaced by a nonarchimedean field, like the field of Puiseux series. In this case, methods from tropical geometry can be applied and are expected to allow one, in generic situations, to reduce semialgebraic problems to combinatorial problems, involving only the nonarchimedean valuations (leading exponents) of the coefficients of the input.

To this purpose, we studied tropical spectrahedra, which are defined as the images by the valuation of nonarchimedean spectrahedra. We establish that they are closed semilinear sets, and that, under a genericity condition, they are described by explicit inequalities expressing the nonnegativity of tropical minors of order 1 and 2. These results are presented in the preprint [52] (now accepted for publication in Disc. Comp. Geom), with further results in the PhD thesis [118].

We showed in [53] that the feasibility problem for a generic tropical spectrahedron is equivalent to solving a stochastic mean payoff game (with perfect information). The complexity of these games is a long-standing open problem. They are not known to be polynomial, however they belong to the class $\mathsf{NP} \cap \mathsf{coNP}$, and they can be solved efficiently in practice. This allows to apply stochastic game algorithms to solve nonarchimedean semidefinite feasibility problems. We obtain in this way both theoretical bounds and a practicable method which solves some large scale instances.

A long-standing problem is to characterize the convex semialgebraic sets that are SDP representable, meaning that they can be represented as the image of a spectrahedron by a (linear) projector. Helton and Nie conjectured that every convex semialgebraic set over the field of real numbers are SDP representable. Recently, [114] disproved this conjecture. In [15], we show, however, that the following result, which may be thought of as a tropical analogue of this conjecture, is true: over a real closed nonarchimedean field of Puiseux series, the convex semialgebraic sets and the projections of spectrahedra have precisely the same images by the nonarchimedean valuation. The proof relies on game theory methods and on our previous results [52] and [53].

In [50] and [118], we exploited the tropical geometry approach to introduce a condition number for stochastic mean payoff games (with perfect information). This condition number is defined as the maximal radius of a ball in Hilbert's projective metric, contained in a primal or dual feasible set. We show that the convergence time of value iteration is governed by this condition number, and derive fixed parameter tractability results.

7.4.2. Tropical polynomial systems and colorful interior of convex bodies

Participants: Marianne Akian, Marin Boyet, Xavier Allamigeon, Stéphane Gaubert.

We studied tropical polynomial systems, with motivations from call center performance evaluation (see Section 7.6.1). We introduced a notion of colorful interior of a family of convex bodies, and showed that the solution of such a polynomial system reduces to linear programming if one knows a vector in the colorful interior of an associated family of Newton polytopes. Further properties of colorful interiors were investigated, as well as the relation between tropical colorful interiors and support vector machines. These results were presented by M. Boyet at the SIAM AG conference in Bern.

7.4.3. Universal approximation theorems by log-sum-exp neural networks

Participant: Stéphane Gaubert.

This is a joint work with Giuseppe Calafiore and Corrado Possieri (Torino).

We establish universal properties of functions by neural networks with log-sum-exp activation functions, first for convex functions [19], and then in general [29]. Some consequences, including approximation by subtraction free rational expressions, are derived.

7.5. Tropical algebra, number theory and directed algebraic topology

7.5.1. An arithmetic site of Connes-Consani type for number fields with narrow class number 1

Participant: Aurélien Sagnier.

In 1995, A. Connes ([65]) gave a spectral interpretation of the zeroes of the Riemann zeta function involving the action of $\mathbb{R}_+^*$ on the sector $X = \mathbb{Q}_+^\times \backslash \mathbb{A}_\mathbb{Q} / \widehat{\mathbb{Z}}^\times$ of the adele class space $\mathbb{A}_\mathbb{Q}/\mathbb{Q}^*$ of the field of rational numbers. In [66], [68], the action of $\mathbb{R}_+^*$ on this sector X was shown to have a natural interpretation in algebraic geometry. This interpretation requires the use of topos theory as well as of the key ingredient of characteristic one namely the semifield $\mathbb{R}_{\max}$ familiar in tropical geometry. The automorphism group of this semifield is naturally isomorphic to $\mathbb{R}_+^*$ and plays the role of the Frobenius. As it turns out, its action on the points of a natural semiringed topos corresponds canonically to the above action on X. This semiringed topos is called the arithmetic site. In my PhD, I extended the construction of the arithmetic site, replacing the field of rational numbers by certain number fields. I considered the simplest complex case, namely that of imaginary quadratic fields on which we assume that the units are not reduced to ± 1 that is when K is either $\mathbb{Q}(\imath)$ or

$\mathbb{Q}(\imath\sqrt{3})$. In particular, during this year, we showed that the semiring of convex polygons introduced for those cases satisfies a sublte arithmetical universal property. These results are presented in the accepted in *Journal of Number Theory* article [111]. In a further work, developed this year, I extended this construction, dealing now with number fields K with narrow class number 1, this generalization will rely on the universal property discovered this year and on the extensive use of Shintani's unit theorem. Here again tropical algebra play a crucial role in the geometrical constructions.

7.5.2. Duality between tropical modules and congruences

Participants: Stéphane Gaubert, Aurélien Sagnier.

In a joint work with Éric Goubault (LIX, École polytechnique), we establish a duality theorem between congruences and modules over tropical semifields.

7.5.3. Directed topological complexity and control

Participant: Aurélien Sagnier.

This is a joint work with Michael Farber and Eric Goubault.

The view we are taking here is that of topological complexity, as defined in [71], adapted to directed topological spaces.

Let us briefly motivate the interest of a directed topological complexity notion. It has been observed that the very important planification problem in robotics boils down to, mathematically speaking, finding a section to the path space fibration $\chi \ : \ PX = X^I \to X \times X$ with $\chi(p) = (p(0), p(1))$. If this section is continuous, then the complexity is the lowest possible (equal to one), otherwise, the minimal number of discontinuities that would encode such a section would be what is called the topological complexity of X. This topological complexity is both understandable algorithmically, and topologically, e.g. as s having a continuous section is equivalent to X being contractible. More generally speaking, the topological complexity is defined as the Schwartz genus of the path space fibration, i.e. is the minimal cardinal of partitions of $X \times X$ into "nice" subspaces F_i such that $s_{F_i} \ : \ F_i \to PX$ is continuous.

This definition perfectly fits the planification problem in robotics where there are no constraints on the actual control that can be applied to the physical apparatus that is supposed to be moved from point a to point b. In many applications, a physical apparatus may have dynamics that can be described as an ordinary differential equation in the state variables $x \in \mathbb{R}^n$ and in time t, parameterized by control parameters $u \in \mathbb{R}^p$, $\dot{x}(t) = f(t, x(t))$. These parameters are generally bounded within some set U, and, not knowing the precise control law (i.e. parameters u as a function of time t) to be applied, the way the controlled system can evolve is as one of the solutions of the differential inclusion $\dot{x}(t) \in F(t, x(t))$ where $F(t, x(t))$ is the set of all $f(t, x(t), u)$ with $u \in U$. Under some classical conditions, this differential inclusion can be proven to have solutions on at least a small interval of time, but we will not discuss this further here. Under the same conditions, the set of solutions of this differential inclusion naturally generates a dspace (a very general structure of directed space, where a preferred subset of paths is singled out, called directed paths, see e.g. [83]). Now, the planification problem in the presence of control constraints equates to finding sections to the analogues to the path space fibration (That would most probably not qualify for being called a fibration in the directed setting) taking a dipath to its end points. This notion is developed in this article, and we introduce a notion of directed homotopy equivalence that has precisely, and in a certain non technical sense, minimally, the right properties with respect to this directed version of topological complexity.

This notion of directed topological complexity also has applications in informatics where a directed space can be used to model the space of all possible executions of a concurrent process (ie when several running programs must share common limited ressources).

In the article [22], after defining the notion of directed topological complexity,this invariant (directed topological complexity) is studied for directed spheres and directed graphs.

7.6. Applications

7.6.1. Performance evaluation of emergency call centers

Participants: Xavier Allamigeon, Marin Boyet, Baptiste Colin, Stéphane Gaubert.

Since 2014, we have been collaborating with Préfecture de Police (Régis Reboul and LcL Stéphane Raclot), more specifically with Brigade de Sapeurs de Pompiers de Paris (BSPP) and Direction de Sécurité de Proximité de l'agglomération parisienne (DSPAP), on the performance evaluation of the new organization (PFAU, "Plate forme d'appels d'urgence") to handle emergency calls to firemen and policemen in the Paris area. We developed analytical models, based on Petri nets with priorities, and fluid limits, see [46], [47], [59]. In 2019, with four students of École polytechnique, Céline Moucer, Julia Escribe, Skandère Sahli and Alban Zammit, we performed case studies, showing the improvement brought by the two level filtering procedure.

Moreover, in 2019, this work was extended to encompass the handling of health emergency calls, with a new collaboration, involving responsibles from the four services of medical emergency aid of Assistance Publique – Hôpitaux de Paris (APHP), i.e., with SAMU75, 92, 93, 94, in the framework of a project led by Dr. Christophe Leroy from APHP. As part of his PhD work, Marin Boyet developed Petri net models capturing the characteristic of the centers (CRRA) handling emergency calls the SAMU, in order to make dimensioning recommendations.

7.6.2. Game theory and optimization methods for decentralized electric systems

Participants: Stéphane Gaubert, Paulin Jacquot.

This work is in collaboration with Nadia Oudjane, Olivier Beaude and Cheng Wan (EDF Lab).

The PhD work of Paulin Jacquot concerns the application of game theory and distributed optimization techniques to the operation of decentralized electric systems, and in particlar to the management of distributed electric consumption flexibilities. We start by adopting the point of view of a centralized operator in charge of the management of flexibilities for several agents. We provide a distributed and privacy-preserving algorithm to compute consumption profiles for agents that are optimal for the operator. In the proposed method, the individual constraints as well as the individual consumption profile of each agent are never revealed to the operator or the other agents [33] [28]. A patent related to this method has been submitted [89].

Then, in a second model, we adopt a more decentralized vision and consider a game theoretic framework for the management of consumption flexibilities. This approach enables, in particular, to take into account the strategic behavior of consumers. Individual objectives are determined by dynamic billing mechanisms, which is motivated by the modeling of congestion effects occurring on time periods receiving a high electricity load from consumers. A relevant class of games in this framework is given by atomic splittable congestion games. We obtain several theoretical results on Nash equilibria for this class of games, and we quantify the efficiency of those equilibria by providing bounds on the price of anarchy. We address the question of the decentralized computation of equilibria in this context by studying the conditions and rates of convergence of the best response and projected gradients algorithms [91], [88].

A fruitful collaboration with Cheng Wan (EDF Lab) led to the third part of this PhD thesis. In this part, we consider an operator dealing with a very large number of players, for which evaluating the equilibria in a congestion game will be difficult. To address this issue, we give approximation results on the equilibria in congestion and aggregative games with a very large number of players, in the presence of coupling constraints. These results, obtained in the framework of variational inequalities and under some monotonicity conditions, can be used to compute an approximate equilibrium, solution of a small dimension problem [32]. In line with the idea of modeling large populations, we consider nonatomic congestion games with coupling constraints, with an infinity of heterogeneous players: these games arise when the characteristics of a population are described by a parametric density function. Under monotonicity hypotheses, we prove that Wardrop equilibria of such games, given as solutions of an infinite dimensional variational inequality, can be approximated by symmetric Wardrop equilibria of auxiliary games, solutions of low dimension variational inequalities. Again, those results can be the basis of tractable methods to compute an approximate Wardrop equilibrium in a nonatomic infinite-type congestion game [33]. Last, in a collaboration with Hélène Le Cadre, Cheng Wan and

Clémence Alasseur, we consider a game model for the study of decentralized peer-to-peer energy exchanges between a community of consumers with renewable production sources. We study the generalized equilibria in this game, which characterize the possible energy trades and associated individual consumptions. We compare the equilibria with the centralized solution minimizing the social cost, and evaluate the efficiency of equilibria through the price of anarchy [23].

Paulin Jacquot defended his PhD on December 5, 2019 at Ecole polytechnique [90].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Decentralized mechanisms of operation of power systems: equilibria and efficiency. Collaboration with Nadia Oudjane and Olivier Beaude from EDF-labs, with the PhD work of Paulin Jacquot (CIFRE PhD), supervised by Stéphane Gaubert.
- Stochastic optimization of multiple flexibilities and energies in micro-grids, collaboration with Wim Van Ackooij, from EDF labs, with the PhD work of Maxime Grangereau (CIFRE PhD), supervised by Emmanuel Gobet (CMAP) and cosupervised by Stéphane Gaubert.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

- Projet ANR JCJC CAPPS ("Combinatorial Analysis of Polytopes and Polyhedral Subdivisions"), responsable Arnau Padrol (IMJ-PRG, Sorbonne Université). Partenaires : IMJ-PRG (Sorbonne Université), Inria Saclay (Tropical), LIGM (Université Paris-Est Marne-la-Vallée), LIF (Université Aix-Marseille), CERMICS (École Nationale des Ponts et Chaussées), LIX (École Polytechnique).

9.1.2. Labex Hadamard

- Projet du Labex Hadamard, intitulé "ALgebraic Methods in gAmes and optimization ALMA", conjoint avec le PGMO, coordonné par E. Tisgaridas (Inria Paris) et X. Allamigeon, faisant intervenir M. Akian et S. Gaubert.

9.1.3. IRS iCODE (Institut pour le Contrôle et la Décision de l'Idex Paris-Saclay)

- White project "New perspectives in the numerical solution of Hamilton-Jacobi-Bellman partial differential equations", coordinated by M. Akian, including S. Gaubert and members of the EPC Commands (Inria Saclay and École polytechnique), UMA (ENSTA), and LMO (Paris-Sud).

9.1.4. Centre des Hautes Études du Ministère de l'Intérieur

- Project "Optimisation de la performance de centres de traitement d'appels d'urgence en cas d'événements planifiés ou imprévus", coordinated by X. Allamigeon, involving M. Boyet, B. Colin and S. Gaubert.

9.2. International Initiatives

9.2.1. Participation in Other International Programs

- Bilateral projects FACCTS, between the University of Chicago (Statistics) – Lek-Heng Lim– and Ecole polytechnique – Stéphane Gaubert– "Tropical geometry of deep learning".

- Math AmSud Project ARGO, “Algebraic Real Geometry and Optimization”, accepted, with CMM (Chile), Univ. Buenos Aires (Argentina), Univ. Fed. Rio and Univ. Fed. Ceara (Brasil), Univ Savoie and CMAP, Ecole polytechnique (France).

9.3. International Research Visitors

9.3.1. Visits of International Scientists

- Oliver Lorscheid, IMPA, Rio (on sabattical at MPI, Bonn), one week in June and 3 days in October, joint invitation with CMLS, Ecole polytechnique.
- Louis Rowen, Bar Ilan University, 3 days in March.
- Sergei Sergeev, Birmingham, 1 week in April.
- Grigorio Malajovich, Univ. Federal, Rio, 1 week in August.
- Armando Gutiérrez, Aalto University, 2 days in February.

9.3.1.1. Internships

- Sarah Vannucci, PhD student, University of Salerno, has been invited for 3 months in the team.

9.3.2. Visits to International Teams

- S. Gaubert
  - Univ. Birmingham, Math and Stats Dep, Jan. 2019 (visiting S. Sergeev)
  - Univ. Bar Ilan, Math Dep, June 2019 (visiting L. Rowen)
  - Univ. Baltimore, Math Dep, Oct. 2019 (visiting A. Sagnier)
- B. Tran
  - U. de Hong Kong, March-April 2019 (2 months, visiting Zheng Qu)
- C. Walsh
  - Univ. Kent, School of Mathematics, Statistics and Actuarial Science, 1 week in November (visiting B. Lemmens and M. Roelands).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- M. Akian is co-chair of the organizing committee of the next 2020 SMAI MODE days, see http://smai-mode2020.inria.fr/.
- Stéphane Gaubert is the coordinator of the Gaspard Monge Program for Optimization, Operations Research and their interactions with data sciences (PGMO), a corporate sponsorhip program, operated by Fondation Mathématique Jacques Hadamard, supported by Criteo, EDF, Orange and Thales, see http://www.fondation-hadamard.fr/fr/pgmo/.

10.1.1.2. Member of the Organizing Committees

- S. Gaubert co-organizes the “Séminaire Parisien d’Optimisation” at Institut Henri Poincaré. https://sites.google.com/site/spoihp/.
- A. Sagnier co-organized the JAMI Workshop, Riemann-Roch theorem in characteristic one and related topics, Univ. Baltimore, Oct. 2019.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- S. Gaubert, Chair of the PGMO days, EDF Lab Paris Saclay, Dec 3-4, 2019. http://www.fondation-hadamard.fr/fr/pgmo/pgmodays

10.1.2.2. Member of the Conference Program Committees

- Marianne Akian is member of the scientific committee of the next 2020 SMAI MODE days, see http://smai-mode2020.inria.fr/.
- S. Gaubert has been a member of the scientific committee of the JAMI Workshop, Riemann-Roch theorem in characteristic one and related topics, Univ. Baltimore, Oct. 2019.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- S. Gaubert is member of the editorial committee of the collection Mathématiques et Applications, SMAI and Springer.
- S. Gaubert is associate editor of Linear and Multilinear Algebra.
- S. Gaubert is associate editor of RAIRO Operations research.
- S. Gaubert is associate editor of the Journal of dynamics and games.

10.1.4. Invited Talks

- M. Akian
  - "Complexity of policy and value iterations for ergodic zero-sum two player games: non-linear Perron-Frobenius methods", Journées annuelles du GDR MOA, Rennes, October 2019.
- S. Gaubert
  - "Condition numbers in nonarchimedean semidefinite programming ...and what they say about stochastic mean payoff games", Tropical Mathematics group of the London Math. Soc, Birmingham, January 2019.
  - "Nonarchimedean convex programming and its relation to mean payoff games", Game theory conference, CIRM, Luminy, June 2019.
- X. Allamigeon
  - "Tropical geometry meets optimization", conference "Network Games, Tropical Geometry, and Quantum Communication", ZIB, Berlin, June 2019

10.1.5. Leadership within the Scientific Community

- See Section 10.1.1.1 (coordination of PGMO).

10.1.6. Research Administration

- M. Akian :
  - Member of the "comité de liaison SMAI-MODE" since June 2015.
  - Member of the laboratory council of CMAP.
  - Member of the scientific board of Inria.
- S. Gaubert :
  - Chairman of the Gaspard Monge Program for Optimization, Operations Research and their interactions with data sciences (PGMO), see 10.1.1.1 for details.
  - Member of the scientific council of CMAP.
- X. Allamigeon:

  - Member of the scientific committee of Inria Saclay – Ile-de-France.
  - Member of the Applied Mathematics Department committee at Ecole Polytechnique.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- M. Akian
  - Course “Markov decision processes: dynamic programming and applications” joint between (3rd year of) ENSTA and M2 “Mathématiques et Applications”, U. Paris Saclay, “Optimization”, and shared with Jean-Philippe Chancelier (ENPC), 15 hours each.
- X. Allamigeon
  - Petites classes et encadrement d’enseignements d’approfondissement de Recherche Opérationnelle en troisième année à l’École Polytechnique (programme d’approfondissement de Mathématiques Appliquées) (niveau M1).
  - Cours du M2 “Optimisation” de l’Université Paris Saclay, cours partagé avec Manuel Ruiz (RTE) et Dominique Quadri (LRI, Université Paris Sud).
  - Co-responsabilité du programme d’approfondissement en mathématiques appliquées (troisième année) à l’École Polytechnique.
- S. Gaubert
  - Course “Systèmes à Événements Discrets”, option MAREVA, ENSMP.
  - Course “Algèbre tropicale pour le contrôle optimal et les jeux” of “Contrôle, Optimisation et Calcul des Variations” (COCV) of M2 “Mathématiques et Applications” of Paris 6 University and École Polytechnique.
  - Lecture of Operations Research, third year of École Polytechnique. The lectures notes were published as a book [60].
- O. Saadi
  - Exercises classes in the framework of a “Monitorat”.
- B. Tran
  - Exercises classes for the first year of Bachelor program of Ecole polytechnique in the framework of a “Monitorat”.

10.2.2. Supervision

- PhD: Paulin Jacquot, registered at Univ. Paris Saclay since November 2016, thesis supervisor: Stéphane Gaubert, cosupervision: Nadia Oujdane, Olivier Beaude (EDF), the defense took place in Dec. 2019.
- PhD in progress: Benoît Tran, registered at Univ Paris-Est Marne La Vallée, since September 2017, thesis supervisor: Jean-Philippe Chancelier (ENPC), cosupervision: Marianne Akian.
- PhD in progress: Maxime Grangereau, registered at Univ. Paris Saclay since Jan 2018, thesis supervisor: Emanuel Gobet, cosupervision: Stéphane Gaubert.
- PhD in progress: Omar Saadi, registered at Univ. Paris Saclay since October 2018, thesis supervisor: Stéphane Gaubert, cosupervision: Marianne Akian.
- PhD in progress: Marin Boyet, registered at Univ. Paris Saclay since October 2018, thesis supervisor: Stéphane Gaubert, cosupervision: Xavier Allamigeon.
- PhD in progress: Maël Forcier, registered at ENPC since September 2019, thesis supervisor: Vincent Leclère, cosupervision Stéphane Gaubert.

10.2.3. Juries

- M. Akian
  - Jury of the 2019 competition for a full professor position in Applied Mathematics at Avignon University.
- S. Gaubert
  - Jury of the 2019 competition for CR positions of Inria Saclay–Île-de-France.
  - Jury of the HdR of Pascale Bendotti (Sorbonne Universités, reviewer, June 2019).
  - Jury of the HdR of Arnau Padrol (Sorbonne Universités, examiner, December 2019).
  - Jury of the PhD thesis of Aiwen Li (Angers, reviewer, Sep 2019).
  - Jury of the PhD thesis of J. Trunk (TU-Berlin, reviewer, Oct 2019),
  - Jury of the PhD thesis of Paul Beaujean (Dauphine, reviewer, Dec. 2019).
  - Jury of the PhD thesis of Paulin Jacquot (Ecole Polytechnique, examiner, Dec. 2019).
  - Jury of the PhD thesis of Arnaud Le Rhun (Ecole Polytechnique, president, Dec. 2019).
- C. Walsh
  - Jury of the PhD thesis of Armando Gutiérrez (Aalto University, pre-examiner, Dec. 2019).

10.3. Conferences, Seminars

- M. Akian
  - "Linear algebra and convexity over symmetrized semirings, hyperfields and systems", SIAM Conference on Applied Algebraic Geometry, Bern, July 2019.
  - "The operator approach to entropy games", International Workshop on Operator Theory and Applications (IWOTA), Lisbonne, July 2019.
  - "Linear algebra and convexity over symmetrized semirings, hyperfields and systems", International Conference on Matrix Analysis and its Applications, MAT TRIAD 2019, Liblice, Czech Republic, September 2019.
  - "A game theory approach to the existence and uniqueness of nonlinear Perron-Frobenius eigenvectors", French-German-Swiss (FGS) conference in Optimization, September 2019, Nice.
  - "De l'ergodicité des jeux à somme nulle à l'existence et l'unicité de vecteurs propres de Perron non linéaires", Workshop "Jeux dynamiques: temps discret, temps continu", Fréjus, October 2019.
  - "Complexity of policy and value iterations for ergodic zero-sum two player games: nonlinear Perron-Frobenius methods", Journées annuelles du GDR MOA, Rennes, October 2019.
- X. Allamigeon
  - Condition numbers of stochastic mean payoff games and what they say about nonarchimedean semidefinite programming, SIAM Conference on Applied Algebraic Geometry, Bern, July 2019.
- M. Boyet
  - The shadow vertex algorithm solves colorful one-versus-all tropical polynomial systems, PGMO Days, Nov. 21, 2018, Palaiseau.
- S. Gaubert
  - Condition numbers in nonarchimedean semidefinite programming ...and what they say about stochastic mean payoff games, Tropical Mathematics group of the London Math. Soc, Birmingham, January 2019.

- Nonarchimedean convex programming and its relation to mean payoff games, Séminaire, Université de Perpignan, March 2019.
- Nonarchimedean convex programming and its relation to mean payoff games, Game theory conference, CIRM, Luminy, June 2019.
- Dynamics of priority: from emergency call centers to tropical polynomial systems, SIAM Conference on Applied Algebraic Geometry, Bern, July 2019.
- Dynamic programming operators over noncommutative spaces: an approach to optimal control of switched systems, International Workshop on Operator Theory and Applications (IWOTA), Lisbonne, July 2019.
- The operator approach to entropy games, French-German-Swiss conference in Optimization, September 2019, Nice.
- “Convexité ambitropicale – ou comment caractériser les rétracts de Shapley”, Workshop "Jeux dynamiques: temps discret, temps continu", Fréjus, October 2019.
- From tropical to ambitropical convexity, JAMI Workshop, Riemann Roch in characteristic one and related topics, Baltimore, October 2019.

- P. Jacquot
  - A Privacy-preserving Disaggregation Algorithm for Non-intrusive Management of Flexible Energy, IEEE 58th Conference on Decision and Control, Nice, France, December 2019.
  - Peer-to-Peer Electricity Market Analysis: From Variational to Generalized Nash Equilibrium, PGMO Days, December 2019, EDF Saclay, France.
  - A Privacy-Preserving Disaggregation Algorithm for Nonconvex Optimization based on Alternate Projections, French-German-Swiss Conference on Optimization (FGS), Nice, France, September 2019.
  - Nonatomic Aggregative Games with Infinitely Many Types, Game Theory Seminar, Insitut Henri Poincaré, Paris, France, October 2019.
  - A Privacy-preserving Method to Optimize Distributed Resource Allocation, Journée de rentrée du CMAP (Invited), Ecole polytechnique, France, October 2019.
- A. Sagnier
  - Talk at University of Strasbourg on «An arithmetic site at the complex place», April 2019
  - Talk at University of Antwerp on «An arithmetic site at the complex place»,

    April 2019
  - Short communication at JAMI 2019 «Riemann-Roch in characteristic one and related topics» on «An arithmetic site at the complex place», Johns Hopkins University, Baltimore, October 2019
- B. Tran
  - “A Min-Plus / SDDP Algorithm for
    Multistage Stochastic Convex Programming”, The XV international conference on stochastic programming (ICSP 2019), Trondheim, July 2019.
  - “A Min-plus-SDDP Algorithm for Multistage Stochastic Convex Programming”, French-German-Swiss Conference on Optimization (FGS), Nice, France, September 2019.
  - “A Min-plus-SDDP Algorithm for Deterministic Multistage Convex Programming”, 58th IEEE Conference on Decision and Control (CDC 2019), Nice, December 2019.
- C. Walsh
  - “For which ordered vector spaces are all order isomorphisms affine linear?”, Advances in the Geometric and Analytic Theory of Convex Cones, Jeju, Korea, May 27–31, 2019.

– “Approximability of convex bodies and volume growth in Hilbert geometries”, Perspectives on convex projective geometry, Sète, France, 24–28 June 2019.
– “For which ordered vector spaces are all order isomorphisms affine linear?”, Positivity X, Pretoria, South Africa, 8–12 July, 2019.
– “For which ordered vector spaces are all order isomorphisms affine linear?”, 30th International Workshop on Operator Theory and its Applications, IWOTA 2019, Lisbon, Portugal, 22–26 July, 2019.

10.4. Popularization

10.4.1. Articles and contents

- The collaboration developed by our team since 2014, with Préfecture de Police, on the performance evaluation of the new organization (PFAU, Plate forme d'appels d'urgences) to handle the calls to the emergency numbers 17-18-112 in the Paris area is described in the following article, published on the Inria web site, https://www.inria.fr/centre/saclay/actualites/gestion-des-appels-d-urgence-une-geometrie-tropicale-pour-les-secours-parisiens

11. Bibliography

Major publications by the team in recent years

[1] M. AKIAN, S. GAUBERT, R. BAPAT.*Non-archimedean valuations of eigenvalues of matrix polynomials*, in "Linear Algebra and its Applications", June 2016, vol. 498, p. 592–627, Also arXiv:1601.00438 [*DOI :* 10.1016/J.LAA.2016.02.036], https://hal.inria.fr/hal-01251803

[2] M. AKIAN, S. GAUBERT, A. GUTERMAN.*Tropical polyhedra are equivalent to mean payoff games*, in "Internat. J. Algebra Comput.", 2012, vol. 22, n^o 1, 1250001, 43, http://dx.doi.org/10.1142/S0218196711006674

[3] M. AKIAN, S. GAUBERT, R. NUSSBAUM.*Uniqueness of the fixed point of nonexpansive semidifferentiable maps*, in "Transactions of the American Mathematical Society", February 2016, vol. 368, n^o 2, Also arXiv:1201.1536 [*DOI :* 10.1090/S0002-9947-2015-06413-7], https://hal.inria.fr/hal-00783682

[4] M. AKIAN, S. GAUBERT, C. WALSH.*The max-plus Martin boundary*, in "Doc. Math.", 2009, vol. 14, p. 195–240

[5] X. ALLAMIGEON, P. BENCHIMOL, S. GAUBERT, M. JOSWIG.*Combinatorial simplex algorithms can solve mean payoff games*, in "SIAM J. Opt.", 2015, vol. 24, n^o 4, p. 2096–2117

[6] X. ALLAMIGEON, P. BENCHIMOL, S. GAUBERT, M. JOSWIG.*Log-barrier interior point methods are not strongly polynomial*, in "SIAM Journal on Applied Algebra and Geometry", 2018, vol. 2, n^o 1, p. 140-178, https://arxiv.org/abs/1708.01544 - This paper supersedes arXiv:1405.4161. 31 pages, 5 figures, 1 table [*DOI :* 10.1137/17M1142132], https://hal.inria.fr/hal-01674959

[7] X. ALLAMIGEON, S. GAUBERT, E. GOUBAULT, S. PUTOT, N. STOTT.*A scalable algebraic method to infer quadratic invariants of switched systems*, in "Proceedings of the International Conference on Embedded Software (EMSOFT)", 2015, Best paper award. The extended version of this conference article appeared in ACM Trans. Embed. Comput. Syst., 15(4):69:1–69:20, September 2016

[8] J. BOLTE, S. GAUBERT, G. VIGERAL.*Definable zero-sum stochastic games*, in "Mathematics of Operations Research", 2014, vol. 40, n$^{\text{o}}$ 1, p. 171–191, Also arXiv:1301.1967, http://dx.doi.org/10.1287/moor.2014.0666

[9] S. FRIEDLAND, S. GAUBERT, L. HAN.*Perron-Frobenius theorem for nonnegative multilinear forms and extensions*, in "Linear Algebra and its Applications", 2013, vol. 438, n$^{\text{o}}$ 2, p. 738–749, This paper was included in a list of “10 Notable Papers from the journal Linear Algebra & Its Applications over the last 50 years” at the occasion of the golden anniversary of the journal, celebrated in 2018., http://dx.doi.org/10.1016/j.laa.2011.02.042

[10] S. GAUBERT, T. LEPOUTRE.*Discrete limit and monotonicity properties of the Floquet eigenvalue in an age structured cell division cycle model*, in "J. Math. Biol.", 2015, http://dx.doi.org/10.1007/s00285-015-0874-3

[11] S. GAUBERT, G. VIGERAL.*A maximin characterization of the escape rate of nonexpansive mappings in metrically convex spaces*, in "Math. Proc. of Cambridge Phil. Soc.", 2012, vol. 152, p. 341–363, Also arXiv:1012.4765, http://dx.doi.org/10.1017/S0305004111000673

[12] C. WALSH.*The horofunction boundary and isometry group of the Hilbert geometry*, in "Handbook of Hilbert Geometry", IRMA Lectures in Mathematics and Theoretical Physics, European Mathematical Society, 2014, vol. 22, https://hal.inria.fr/hal-00782827

Publications of the year

Articles in International Peer-Reviewed Journal

[13] M. AKIAN, S. GAUBERT, J. GRAND-CLÉMENT, J. GUILLAUD.*The operator approach to entropy games*, in "Theory of Computing Systems", 2019, https://arxiv.org/abs/1904.05151 , forthcoming [*DOI :* 10.1007/S00224-019-09925-Z], https://hal.archives-ouvertes.fr/hal-02143807

[14] M. AKIAN, S. GAUBERT, A. HOCHART.*A game theory approach to the existence and uniqueness of nonlinear Perron-Frobenius eigenvectors*, in "Discrete and Continuous Dynamical Systems - Series A", 2020, vol. 40, p. 207–231, https://arxiv.org/abs/1812.09871 [*DOI :* 10.3934/DCDS.2020009], https://hal.inria.fr/hal-01967495

[15] X. ALLAMIGEON, S. GAUBERT, M. SKOMRA.*The tropical analogue of the Helton–Nie conjecture is true*, in "Journal of Symbolic Computation", March 2019, vol. 91, p. 129-148, https://arxiv.org/abs/1801.02089 - The present results were announced in the conference MEGA2017 (INTERNATIONAL CONFERENCE ON EFFECTIVE METHODS IN ALGEBRAIC GEOMETRY, Université Nice Sophia Antipolis, June 2017, https://mega2017.inria.fr/) [*DOI :* 10.1016/J.JSC.2018.06.017], https://hal.inria.fr/hal-01674497

[16] X. ALLAMIGEON, R. D. KATZ.*A Formalization of Convex Polyhedra Based on the Simplex Method*, in "Journal of Automated Reasoning", August 2019, vol. 63, n$^{\text{o}}$ 2, p. 323–345 [*DOI :* 10.1007/S10817-018-9477-1], https://hal.archives-ouvertes.fr/hal-01967575

[17] R. BHATIA, S. GAUBERT, T. JAIN.*Matrix versions of the Hellinger distance*, in "Letters in Mathematical Physics", 2019, vol. 109, p. 1777-1804, https://arxiv.org/abs/1901.01378 [*DOI :* 10.1007/S11005-019-01156-0], https://hal.inria.fr/hal-01973935

[18] V. BOEUF, P. ROBERT.*A Stochastic Analysis of a Network with Two Levels of Service*, in "Queueing Systems", August 2019, vol. 92, n$^{\text{o}}$ 3-4, 30, https://arxiv.org/abs/1708.09590 , forthcoming [*DOI :* 10.1007/S11134-019-09617-Y], https://hal.archives-ouvertes.fr/hal-01583704

[19] G. C. CALAFIORE, S. GAUBERT, C. POSSIERI.*Log-sum-exp neural networks and posynomial models for convex and log-log-convex data*, in "IEEE Transactions on Neural Networks and Learning Systems", 2019, https://arxiv.org/abs/1806.07850 [*DOI :* 10.1109/TNNLS.2019.2910417], https://hal.inria.fr/hal-01969634

[20] S. GAUBERT, M. MACCAIG.*Approximating the Volume of Tropical Polytopes is Difficult*, in "International Journal of Algebra and Computation", 2019, vol. 29, n^{o} 02, p. 357–389, https://arxiv.org/abs/1706.06467 [*DOI :* 10.1142/S0218196718500686], https://hal.inria.fr/hal-01675715

[21] S. GAUBERT, N. STOTT.*A convergent hierarchy of non-linear eigenproblems to compute the joint spectral radius of nonnegative matrices*, in "Mathematical Control and Related Fields", 2019, https://arxiv.org/abs/1805.03284 - 18 pages [*DOI :* 10.3934/MCRF.2020011], https://hal.inria.fr/hal-01967552

[22] E. GOUBAULT, A. SAGNIER, M. FÄRBER.*Directed topological complexity*, in "Journal of Applied and Computational Topology", August 2019, https://hal.archives-ouvertes.fr/hal-02434377

[23] H. LE CADRE, P. JACQUOT, C. WAN, C. ALASSEUR.*Peer-to-Peer Electricity Market Analysis: From Variational to Generalized Nash Equilibrium*, in "European Journal of Operational Research", 2019, https://arxiv.org/abs/1812.02301 , forthcoming [*DOI :* 10.1016/J.EJOR.2019.09.035], https://hal.archives-ouvertes.fr/hal-01944644

International Conferences with Proceedings

[24] M. AKIAN, J.-P. CHANCELIER, B. TRAN.*A Min-plus-SDDP Algorithm for Deterministic Multistage Convex Programming*, in "58th IEEE Conference on Decision and Control", Nice, France, December 2019, https://hal.inria.fr/hal-02436343

[25] M. AKIAN, S. GAUBERT, Z. QU, O. SAADI.*Solving Ergodic Markov Decision Processes and Perfect Information Zero-sum Stochastic Games by Variance Reduced Deflated Value Iteration*, in "CDC 2019 - 58th IEEE Conference on Decision and Control", Nice, France, Proc. of the 58th IEEE Conference on Decision and Control, December 2019, https://arxiv.org/abs/1909.06185 , https://hal.inria.fr/hal-02423846

[26] P. JACQUOT, O. BEAUDE, P. BENCHIMOL, S. GAUBERT, N. OUDJANE.*A Privacy-preserving Disaggregation Algorithm for Non-intrusive Management of Flexible Energy*, in "CDC 2019 - 58th IEEE Conference on Decision and Control", Nice, France, Proceedings of the 58th IEEE Conference on Decision and Control, IEEE, December 2019, https://arxiv.org/abs/1903.03053 , https://hal.archives-ouvertes.fr/hal-02150209

Scientific Books (or Scientific Book chapters)

[27] M. AKIAN, E. FODJO.*Probabilistic max-plus schemes for solving Hamilton-Jacobi-Bellman equations*, in "Numerical Methods for Optimal Control Problems", M. FALCONE, R. FERRETTI, L. GRUNE, W. MCENEANEY (editors), INDAM Series, Springer, February 2019, vol. 29, p. 183–209, https://arxiv.org/abs/1801.01780 , https://hal.inria.fr/hal-01675068

Other Publications

[28] O. BEAUDE, P. BENCHIMOL, S. GAUBERT, P. JACQUOT, N. OUDJANE.*A Privacy-preserving Method to Optimize Distributed Resource Allocation*, August 2019, https://arxiv.org/abs/1908.03080 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02262271

[29] G. C. CALAFIORE, S. GAUBERT, C. POSSIERI.*A Universal Approximation Result for Difference of log-sum-exp Neural Networks*, December 2019, https://arxiv.org/abs/1905.08503 - working paper or preprint, https://hal.inria.fr/hal-02423871

[30] S. FRIEDLAND, S. GAUBERT.*Spectral inequalities for nonnegative tensors and their tropical analogues*, September 2019, https://arxiv.org/abs/1804.00204 - working paper or preprint, https://hal.inria.fr/hal-01969021

[31] S. GAUBERT, A. NIV.*Tropical planar networks*, December 2019, https://arxiv.org/abs/1910.12934 - working paper or preprint, https://hal.inria.fr/hal-02423904

[32] P. JACQUOT, C. WAN, O. BEAUDE, N. OUDJANE.*Efficient Estimation of Equilibria in Large Aggregative Games with Coupling Constraints*, November 2019, https://arxiv.org/abs/1911.10571 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01904546

[33] P. JACQUOT, C. WAN.*Nonatomic Aggregative Games with Infinitely Many Types*, June 2019, https://arxiv.org/abs/1906.01986 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02146294

[34] C. WALSH.*Order isomorphisms of complete order-unit spaces*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02425988

References in notes

[35] M. AKIAN, J.-P. CHANCELIER, B. TRAN.*A stochastic algorithm for deterministic multistage optimization problems*, December 2018, https://arxiv.org/abs/1810.12870 - working paper or preprint, https://hal.inria.fr/hal-01964189

[36] M. AKIAN, E. FODJO.*A probabilistic max-plus numerical method for solving stochastic control problems*, in "55th Conference on Decision and Control (CDC 2016)", Las Vegas, United States, December 2016, Also arXiv:1605.02816, https://hal.inria.fr/hal-01425344

[37] M. AKIAN, E. FODJO.*From a monotone probabilistic scheme to a probabilistic max-plus algorithm for solving Hamilton-Jacobi-Bellman equations*, in "Hamilton-Jacobi-Bellman Equations: Numerical Methods and Applications in Optimal Control", D. KALISE, K. KUNISCH, Z. RAO (editors), De Gruyter, August 2018, https://arxiv.org/abs/1709.09049 , https://hal.inria.fr/hal-01675067

[38] M. AKIAN, S. GAUBERT.*Spectral theorem for convex monotone homogeneous maps, and ergodic control*, in "Nonlinear Anal.", 2003, vol. 52, n^{o} 2, p. 637–679, http://dx.doi.org/10.1016/S0362-546X(02)00170-0

[39] M. AKIAN, S. GAUBERT.*Policy iteration for perfect information stochastic mean payoff games with bounded first return times is strongly polynomial*, 2013, Preprint arXiv:1310.4953, 17 pages, http://hal.inria.fr/hal-00881207

[40] M. AKIAN, S. GAUBERT, A. GUTERMAN.*Tropical polyhedra are equivalent to mean payoff games*, in "Internat. J. Algebra Comput.", 2012, vol. 22, n^{o} 1, 1250001, 43 [*DOI :* 10.1142/S0218196711006674], http://arxiv.org/abs/0912.2462

[41] M. AKIAN, S. GAUBERT, A. HOCHART.*Generic uniqueness of the bias vector of finite stochastic games with perfect information*, in "Journal of Mathematical Analysis and Applications", 2018, vol. 457, p. 1038-1064, https://arxiv.org/abs/1610.09651 [*DOI :* 10.1016/J.JMAA.2017.07.017], https://hal.inria.fr/hal-01425543

[42] M. AKIAN, S. GAUBERT, M. SHARIFY.*Log-majorization of the moduli of the eigenvalues of a matrix polynomial by tropical roots*, in "Linear Algebra and its Applications", 2017, Also arXiv:1304.2967 [*DOI :* 10.1016/J.LAA.2016.11.004], https://hal.inria.fr/hal-00881196

[43] X. ALLAMIGEON, P. BENCHIMOL, S. GAUBERT.*The tropical shadow-vertex algorithm solves mean payoff games in polynomial time on average*, in "ICALP 2014", Copenhagen, France, J. ESPARZA, P. FRAIGNIAUD, T. HUSFELDT, E. KOUTSOUPIAS (editors), 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I, Springer, July 2014, vol. 8572, 12 [*DOI :* 10.1007/978-3-662-43948-7_8], https://hal.inria.fr/hal-01096447

[44] X. ALLAMIGEON, P. BENCHIMOL, S. GAUBERT, M. JOSWIG.*Long and winding central paths*, May 2014, Preprint arXiv:1405.4161, v2 May 2015, https://hal.inria.fr/hal-01096452

[45] X. ALLAMIGEON, P. BENCHIMOL, S. GAUBERT, M. JOSWIG.*Log-barrier interior point methods are not strongly polynomial*, in "SIAM Journal on Applied Algebra and Geometry", 2018, vol. 2, n^{o} 1, p. 140-178, https://arxiv.org/abs/1708.01544 - This paper supersedes arXiv:1405.4161. 31 pages, 5 figures, 1 table [*DOI :* 10.1137/17M1142132], https://hal.inria.fr/hal-01674959

[46] X. ALLAMIGEON, V. BOEUF, S. GAUBERT.*Performance evaluation of an emergency call center: tropical polynomial systems applied to timed Petri nets*, in "13th International Conference, Formal Modeling and Analysis of Timed Systems (FORMATS 2015)", Madrid, Spain, Formal Modeling and Analysis of Timed Systems, Springer, September 2015, vol. 9268 [*DOI :* 10.1007/978-3-319-22975-1_2], https://hal.inria.fr/hal-01248814

[47] X. ALLAMIGEON, V. BOEUF, S. GAUBERT.*Stationary solutions of discrete and continuous Petri nets with priorities*, in "Performance Evaluation", August 2017, vol. 113, p. 1 - 12, https://arxiv.org/abs/1612.07661 [*DOI :* 10.1016/J.PEVA.2017.04.007], https://hal.inria.fr/hal-01674492

[48] X. ALLAMIGEON, S. GAUBERT, É. GOUBAULT.*Inferring Min and Max Invariants Using Max-plus Polyhedra*, in "Proceedings of the 15th International Static Analysis Symposium (SAS'08)", Valencia, Spain, LNCS, Springer, 2008, vol. 5079, p. 189–204, http://dx.doi.org/10.1007/978-3-540-69166-2_13

[49] X. ALLAMIGEON, S. GAUBERT, E. GOUBAULT.*Computing the Vertices of Tropical Polyhedra using Directed Hypergraphs*, in "Discrete Comp. Geom.", 2012, Published on line [*DOI :* 10.1007/S00454-012-9469-6], http://fr.arxiv.org/abs/0904.3436v3

[50] X. ALLAMIGEON, S. GAUBERT, R. KATZ, M. SKOMRA.*Condition numbers of stochastic mean payoff games and what they say about nonarchimedean semidefinite programming*, in "23rd International Symposium on Mathematical Theory of Networks and Systems", Hong-Kong, France, July 2018, https://arxiv.org/abs/1802.07712 - 14 pages, 2 figures, https://hal.inria.fr/hal-01967555

[51] X. ALLAMIGEON, S. GAUBERT, M. SKOMRA.*Solving Generic Nonarchimedean Semidefinite Programs Using Stochastic Game Algorithms*, in "ISSAC '16: International Symposium on Symbolic and Algebraic Computation", Waterloo, France, Proceedings of the ACM on International Symposium on Symbolic and Algebraic

Computation (ISSAC'16), ACM, July 2016, Also arXiv:1603.06916 [*DOI :* 10.1145/2930889.2930935], https://hal.inria.fr/hal-01422638

[52] X. ALLAMIGEON, S. GAUBERT, M. SKOMRA.*Tropical spectrahedra*, October 2016, arXiv:1610.06746, https://hal.inria.fr/hal-01422639

[53] X. ALLAMIGEON, S. GAUBERT, M. SKOMRA.*Solving generic nonarchimedean semidefinite programs using stochastic game algorithms*, in "Journal of Symbolic Computation", 2018, vol. 85, p. 25-54, An abridged version of this article appeared in the proceedings of ISSAC 2016 [*DOI :* 10.1016/J.JSC.2017.07.002], https://hal.inria.fr/hal-01674494

[54] X. ALLAMIGEON, R. D. KATZ.*A Formalization of Convex Polyhedra Based on the Simplex Method*, in "Journal of Automated Reasoning", August 2018, https://hal.archives-ouvertes.fr/hal-01967575

[55] P. ANDY, W. FAISAL, F. BONNANS.*MIDAS: A Mixed Integer Dynamic Approximation Scheme*, Inria, 2016, https://hal.inria.fr/hal-01401950

[56] F. BACCELLI, G. COHEN, G.-J. OLSDER, J.-P. QUADRAT.*Synchronization and linearity: an algebra for discrete event systems*, Wiley, 1992

[57] G. BARLES, S. MIRRAHIMI, B. PERTHAME.*Concentration in Lotka-Volterra parabolic or integral equations: a general convergence result*, in "Methods Appl. Anal.", 2009, vol. 16, n^{o} 3, p. 321–340, http://dx.doi.org/10.4310/MAA.2009.v16.n3.a4

[58] R. BHATIA, S. GAUBERT, T. JAIN.*Matrix versions of the Hellinger distance*, 2019, To appear in Letters in Math. Physics

[59] V. BOEUF, P. ROBERT.*A Stochastic Analysis of a Network with Two Levels of Service*, August 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01583704

[60] F. BONNANS, S. GAUBERT.*Recherche opérationnelle. Aspects mathématiques et applications*, Ellipse, March 2016, 391, https://hal.inria.fr/hal-01422645

[61] P. BUTKOVIČ.*Max-algebra: the linear algebra of combinatorics?*, in "Linear Algebra and its applications", 2003, vol. 367, p. 313–335

[62] P. BUTKOVIČ.*Max-linear systems: theory and algorithms*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London, 2010, xviii+272, http://dx.doi.org/10.1007/978-1-84996-299-5

[63] J. COCHET-TERRASSON, G. COHEN, S. GAUBERT, M. MC GETTRICK, J.-P. QUADRAT.*Numerical computation of spectral elements in max-plus algebra*, in "Proc. of the IFAC Conference on System Structure and Control", Nantes, July 1998

[64] G. COHEN, S. GAUBERT, J.-P. QUADRAT.*Max-plus algebra and system theory: where we are and where to go now*, in "Annual Reviews in Control", 1999, vol. 23, p. 207–219

[65] A. CONNES.*Trace formula in noncommutative geometry and the zeros of the Riemann zeta function*, in "Selecta Math. (N.S.)", 1999, vol. 5, n^{o} 1, p. 29–106

[66] A. CONNES, C. CONSANI.*The Arithmetic Site*, in "Comptes Rendus Mathématiques", 2014, vol. Ser. I 352, p. 971–975

[67] A. CONNES, C. CONSANI.*Geometry of the arithmetic site*, 2015, arXiv:1502.05580

[68] A. CONNES, C. CONSANI.*Geometry of the arithmetic site*, in "Adv. Math.", 2016, vol. 291, p. 274–329

[69] P. COUSOT, R. COUSOT.*Abstract Interpretation: A unified lattice model for static analysis of programs by construction of approximations of fixed points*, in "Principles of Programming Languages 4", 1977, p. 238–252

[70] A. FAHIM, N. TOUZI, X. WARIN.*A probabilistic numerical method for fully nonlinear parabolic PDEs*, in "Ann. Appl. Probab.", 2011, vol. 21, n° 4, p. 1322–1364, http://dx.doi.org/10.1214/10-AAP723

[71] M. FARBER.*Invitation to Topological Robotics*, Zurich lectures in advanced mathematics, European Mathematical Society, 2008

[72] A. FATHI, A. SICONOLFI.*Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation*, in "Invent. Math.", 2004, vol. 155, n° 2, p. 363–388, http://dx.doi.org/10.1007/s00222-003-0323-6

[73] O. FERCOQ, M. AKIAN, M. BOUHTOU, S. GAUBERT.*Ergodic control and polyhedral approaches to PageRank optimization*, in "IEEE Trans. Automat. Control", 2013, vol. 58, n° 1, p. 134–148, http://dx.doi.org/10.1109/TAC.2012.2226103

[74] W. FLEMING, W. MCENEANEY.*A max-plus based algorithm for an HJB equation of non-linear filtering*, in "SIAM J. Control and Opt.", 2000, p. 683–710

[75] S. FOMIN, A. ZELEVINSKY.*Cluster algebras. I. Foundations*, in "J. Amer. Math. Soc.", 2002, vol. 15, n° 2, p. 497–529, http://arxiv.org/abs/math.RT/0104151

[76] S. GAUBERT, E. GOUBAULT, A. TALY, S. ZENNOU.*Static Analysis by Policy Iteration in Relational Domains*, in "Proceedings of the Proc. of the 16th European Symposium on Programming (ESOP'07)", Braga (Portugal), LNCS, Springer, 2007, vol. 4421, p. 237–252, http://dx.doi.org/10.1007/978-3-540-71316-6_17

[77] S. GAUBERT, J. GUNAWARDENA.*The Perron-Frobenius Theorem for Homogeneous, Monotone Functions*, in "Trans. of AMS", 2004, vol. 356, n° 12, p. 4931-4950, http://www.ams.org/tran/2004-356-12/S0002-9947-04-03470-1/home.html

[78] S. GAUBERT, W. MCENEANEY, Z. QU.*Curse of dimensionality reduction in max-plus based approximation methods: theoretical estimates and improved pruning algorithms*, in "Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC 11)", Orlando, FL, USA, December 2011, p. 1054-1061, http://arxiv.org/abs/1109.5241

[79] S. GAUBERT, A. NIV.*Tropical totally positive matrices*, in "Journal of Algebra", December 2018, vol. 515, p. 511-544, https://arxiv.org/abs/1606.00238 - arXiv:1606.00238 [*DOI :* 10.1016/J.JALGEBRA.2018.07.005], https://hal.inria.fr/hal-01423747

[80] S. GAUBERT, M. SHARIFY.*Tropical scaling of polynomial matrices*, in "Positive systems", Berlin, Lecture Notes in Control and Inform. Sci., Springer, 2009, vol. 389, p. 291–303, http://dx.doi.org/10.1007/978-3-642-02894-6_28

[81] T. M. GAWLITZA, H. SEIDL, A. ADJÉ, S. GAUBERT, E. GOUBAULT.*Abstract interpretation meets convex optimization*, in "J. Symbolic Comput.", 2012, vol. 47, n^{o} 12, p. 1416–1446, Special issue on Invariant generation and reasoning about loops, http://dx.doi.org/10.1016/j.jsc.2011.12.048

[82] I. M. GELFAND, M. M. KAPRANOV, A. V. ZELEVINSKY.*Discriminants, resultants and multidimensional determinants*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2008, x+523, Reprint of the 1994 edition

[83] M. GRANDIS.*Directed Algebraic Topology, Models of non-reversible worlds*, Cambridge University Press, 2009

[84] S. HAMMARLING, C. J. MUNRO, F. TISSEUR.*An algorithm for the complete solution of quadratic eigenvalue problems*, in "ACM Trans. Math. Software", 2013, vol. 39, n^{o} 3, Art. 18, 19, http://dx.doi.org/10.1145/2450153.2450156

[85] B. HEIDERGOTT, G. J. OLSDER, J. VAN DER WOUDE.*Max Plus at Work: Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications*, Princeton, 2005

[86] H. ISHII, H. MITAKE.*Representation formulas for solutions of Hamilton-Jacobi equations with convex Hamiltonians*, in "Indiana Univ. Math. J.", 2007, vol. 56, n^{o} 5, p. 2159–2183, http://dx.doi.org/10.1512/iumj.2007.56.3048

[87] I. ITENBERG, G. MIKHALKIN, E. SHUSTIN.*Tropical algebraic geometry*, Oberwolfach Seminars, Birkhäuser Verlag, Basel, 2007, vol. 35, viii+103

[88] P. JACQUOT, O. BEAUDE, S. GAUBERT, N. OUDJANE.*Demand Side Management in the Smart Grid: an Efficiency and Fairness Tradeoff*, in "7th IEEE International Conference on Innovative Smart Grid Technologies", Torino, France, August 2017, https://arxiv.org/abs/1711.11129 , https://hal.inria.fr/hal-01675658

[89] P. JACQUOT, S. GAUBERT, N. OUDJANE, O. BEAUDE, P. BENCHIMOL.*Procédé de gestion décentralisée de consommation électrique non-intrusif*, EDF and Inria, 2018, French Patent, FR1872553, filed to INPI on 7 Dec. 2018

[90] P. JACQUOT.*Game theory and Optimization Methods for Decentralized Electric Systems*, 2019

[91] P. JACQUOT, C. WAN.*Nonsmooth Aggregative Games with Coupling Constraints and Infinitely Many Classes of Players*, October 2018, https://arxiv.org/abs/1806.06230 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01904527

[92] H. KAISE, W. M. MCENEANEY.*Idempotent expansions for continuous-time stochastic control: compact control space*, in "Proceedings of the 49th IEEE Conference on Decision and Control", Atlanta, Dec. 2010

[93] V. KOLOKOLTSOV, V. MASLOV.*Idempotent analysis and applications*, Kluwer Acad. Publisher, 1997

[94] B. LEMMENS, R. NUSSBAUM.*Nonlinear Perron-Frobenius theory*, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2012, vol. 189, xii+323, http://dx.doi.org/10.1017/CBO9781139026079

[95] Q. LU, M. MADSEN, M. MILATA, S. RAVN, U. FAHRENBERG, K. G. LARSEN.*Reachability Analysis for Timed Automata using Max-Plus Algebra*, in "J. Logic Alg. Prog.", 2012, vol. 81, n° 3, p. 298-313

[96] V. MASLOV.*Méthodes Operatorielles*, Edition Mir, Moscou, 1987

[97] W. MCENEANEY, A. DESHPANDE, S. GAUBERT.*Curse-of-Complexity Attenuation in the Curse-of-Dimensionality-Free Method for HJB PDEs*, in "Proc. of the 2008 American Control Conference", Seattle, Washington, USA, June 2008

[98] W. M. MCENEANEY, H. KAISE, S. H. HAN.*Idempotent Method for Continuous-time Stochastic Control and Complexity Attenuation*, in "Proceedings of the 18th IFAC World Congress, 2011", Milano, Italie, 2011, p. 3216-3221

[99] W. M. MCENEANEY.*Max-plus methods for nonlinear control and estimation*, Systems & Control: Foundations & Applications, Birkhäuser Boston Inc., Boston, MA, 2006, xiv+241

[100] W. M. MCENEANEY.*A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs*, in "SIAM J. Control Optim.", 2007, vol. 46, n° 4, p. 1239–1276, http://dx.doi.org/10.1137/040610830

[101] J.-F. MERTENS, S. SORIN, S. ZAMIR.*Repeated Games*, Cambridge, 2015

[102] G. MIKHALKIN.*Enumerative tropical algebraic geometry in $\mathbb{R}^2$*, in "J. Amer. Math. Soc.", 2005, vol. 18, n° 2, p. 313–377, http://dx.doi.org/10.1090/S0894-0347-05-00477-7

[103] R. H. MÖHRING, M. SKUTELLA, F. STORK.*Scheduling with AND/OR precedence constraints*, in "SIAM J. Comput.", 2004, vol. 33, n° 2, p. 393–415, http://dx.doi.org/10.1137/S009753970037727X

[104] A. PAPADOPOULOS.*Metric spaces, convexity and non-positive curvature*, IRMA Lectures in Mathematics and Theoretical Physics, Second, European Mathematical Society (EMS), Zürich, 2014, vol. 6, xii+309, http://dx.doi.org/10.4171/132

[105] M. V. F. PEREIRA, L. M. V. G. PINTO.*Multi-stage stochastic optimization applied to energy planning*, in "Math. Programming", 1991, vol. 52, n° 2, Ser. B, p. 359–375, http://dx.doi.org/10.1007/BF01582895

[106] G. C. PFLUG, A. PICHLER.*A Distance For Multistage Stochastic Optimization Models*, in "SIAM Journal on Optimization", January 2012, vol. 22, n° 1, p. 1-23, http://dx.doi.org/10.1137/110825054

[107] J.-E. PIN.*Tropical Semirings*, in "Idempotency", J. GUNAWARDENA (editor), Publications of the Isaac Newton Institute, Cambridge University Press, 1998, vol. 11, p. 50–69

[108] Z. QU.*Théorie de Perron-Frobenius non linéaire et méthodes numériques max-plus pour la résolution d'équations d'Hamilton-Jacobi*, Ecole Polytechnique X, October 2013, http://hal.inria.fr/pastel-00927122

[109] D. REEB, M. J. KASTORYANO, M. M. WOLF.*Hilbert's projective metric in quantum information theory*, in "J. Math. Phys.", 2011, vol. 52, n° 8, 082201, 33, http://dx.doi.org/10.1063/1.3615729

[110] J. RICHTER-GEBERT, B. STURMFELS, T. THEOBALD.*First steps in tropical geometry*, in "Idempotent mathematics and mathematical physics", Providence, RI, Contemp. Math., Amer. Math. Soc., 2005, vol. 377, p. 289–317

[111] A. SAGNIER.*An arithmetic site at the complex place*, in "Journal of Number Theory", 2019, published on line

[112] G. SAGNOL, S. GAUBERT, M. BOUHTOU.*Optimal monitoring on large networks by Successive c-Optimal Designs*, in "Proceedings of the 22nd international teletraffic congress (ITC22), Amsterdam, The Netherlands, September", IEEE, 2010, http://dx.doi.org/10.1109/ITC.2010.5608717

[113] S. SANKARANARAYANAN, H. SIPMA, Z. MANNA.*Scalable Analysis of Linear Systems using Mathematical Programming*, in "VMCAI", LNCS, 2005, vol. 3385

[114] C. SCHEIDERER.*Semidefinitely representable convex sets*, 2016, arXiv:1612.07048v2

[115] R. SEPULCHRE, A. SARLETTE, P. ROUCHON.*Consensus in noncommutative spaces*, in "Proceedings of the 49th IEEE Conference on Decision and Control", Atlanta, USA, Dec 2010, p. 6596-6601, http://dx.doi.org/10.1109/CDC.2010.5717072

[116] A. SIDFORD, M. WANG, X. WU, Y. YE.*Variance reduced value iteration and faster algorithms for solving Markov decision processes*, in "Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms", SIAM, Philadelphia, PA, 2018, p. 770–787, https://doi.org/10.1137/1.9781611975031.50

[117] I. SIMON.*Limited subsets of a free monoid*, in "Proc. 19th Annual Symposium on Foundations of Computer Science", Piscataway, NJ, 1978, p. 143–150

[118] M. SKOMRA.*Spectraèdres tropicaux : application à la programmation semi-définie et aux jeux à paiement moyen*, Université Paris-Saclay, December 2018, https://pastel.archives-ouvertes.fr/tel-01958741

[119] H. L. SMITH.*Monotone dynamical systems*, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1995, vol. 41, x+174, An introduction to the theory of competitive and cooperative systems

[120] N. M. TRAN, J. YU.*Product-Mix Auctions and Tropical Geometry*, 2015

[121] O. VIRO.*Dequantization of real algebraic geometry on logarithmic paper*, in "European Congress of Mathematics, Vol. I (Barcelona, 2000)", Basel, Progr. Math., Birkhäuser, 2001, vol. 201, p. 135–146

[122] O. VIRO.*Dequantization of real algebraic geometry on logarithmic paper*, in "European Congress of Mathematics, Vol. I (Barcelona, 2000)", Basel, Progr. Math., Birkhäuser, 2001, vol. 201, p. 135–146, http://arxiv.org/abs/math.AG/0005163

Project-Team XPOP

Statistical modelling for life sciences

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

IN PARTNERSHIP WITH:
Ecole Polytechnique

RESEARCH CENTER
Saclay - Île-de-France

THEME
Modeling and Control for Life Sciences

Table of contents

Project-Team XPOP

Creation of the Team: 2016 January 01, updated into Project-Team: 2017 July 01

Keywords:

Computer Science and Digital Science:

A3.1.1. - Modeling, representation
A3.2.3. - Inference
A3.3. - Data and knowledge analysis
A3.3.1. - On-line analytical processing
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A3.4.8. - Deep learning
A5.9.2. - Estimation, modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.3.3. - Data processing
A6.3.5. - Uncertainty Quantification

Other Research Topics and Application Domains:

B1.1.4. - Genetics and genomics
B1.1.7. - Bioinformatics
B1.1.10. - Systems and synthetic biology
B2.2.3. - Cancer
B2.2.4. - Infectious diseases, Virology
B2.4.1. - Pharmaco kinetics and dynamics
B9.1.1. - E-learning, MOOC

1. Team, Visitors, External Collaborators

Research Scientists

Marc Lavielle [Team leader, Inria, Senior Researcher]
Erwan Le Pennec [École polytechnique, Researcher]
Laetitia Le [APHP - Ecole polytechnique, Researcher, until Oct 2019]

Faculty Members

Julie Josse [École polytechnique, Associate Professor]
Eric Moulines [École polytechnique, Professor]

Technical Staff

Yao Xu [Inria, Engineer, until Jun 2019]

PhD Students

Nicolas Brosse [École polytechnique, PhD Student, until Aug 2019]
Wei Jiang [École polytechnique, PhD Student]
Mohammed Karimi [Inria, PhD Student, until Sep 2019]
Genevieve Robin [École polytechnique, PhD Student, until Aug 2019]
Marine Zulian [Dassault Systemes, PhD Student, granted by CIFRE]

Post-Doctoral Fellows

Angie Pineda [École polytechnique, Post-Doctoral Fellow]
Tom Rohmer [Inria, Post-Doctoral Fellow, until Aug 2019]

Visiting Scientist

Ricardo Rios [Universidad Central de Venezuela, Sep 2019]

Administrative Assistants

Ines Dumontier [Inria, until May 2019]
Hanadi Dib [Inria, from May 2019]

2. Overall Objectives

2.1. Developing sound, useful and usable methods

The main objective of XPOP is to develop new sound and rigorous methods for statistical modeling in the field of biology and life sciences. These methods for modeling include statistical methods of estimation, model diagnostics, model building and model selection as well as methods for numerical models (systems of ordinary and partial differential equations). Historically, the key area where these methods have been used is population pharmacokinetics. However, the framework is currently being extended to sophisticated numerical models in the contexts of viral dynamics, glucose-insulin processes, tumor growth, precision medicine, spectrometry, intracellular processes, etc.

Furthermore, an important aim of XPOP is to transfer the methods developed into software packages so that they can be used in everyday practice.

2.2. Combining numerical, statistical and stochastic components of a model

Mathematical models that characterize complex biological phenomena are defined by systems of ordinary differential equations when dealing with dynamical systems that evolve with respect to time, or by partial differential equations when there is a spatial component in the model. Also, it is sometimes useful to integrate a stochastic aspect into the dynamical system in order to model stochastic intra-individual variability.

In order to use such methods, we must deal with complex numerical difficulties, generally related to resolving the systems of differential equations. Furthermore, to be able to check the quality of a model (i.e. its descriptive and predictive performances), we require data. The statistical aspect of the model is thus critical in how it takes into account different sources of variability and uncertainty, especially when data come from several individuals and we are interested in characterizing the inter-subject variability. Here, the tools of reference are mixed-effects models.

Confronted with such complex modeling problems, one of the goals of XPOP is to show the importance of combining numerical, statistical and stochastic approaches.

2.3. Developing future standards

Linear mixed-effects models have been well-used in statistics for a long time. They are a classical approach, essentially relying on matrix calculations in Gaussian models. Whereas a solid theoretical base has been developed for such models, *nonlinear* mixed-effects models (NLMEM) have received much less attention in the statistics community, even though they have been applied to many domains of interest. It has thus been the users of these models, such as pharmacometricians, who have taken them and developed methods, without really looking to develop a clean theoretical framework or understand the mathematical properties of the methods. This is why a standard estimation method in NLMEM is to linearize the model, and few people have been interested in understanding the properties of estimators obtained in this way.

Statisticians and pharmacometricians frequently realize the need to create bridges between these two communities. We are entirely convinced that this requires the development of new standards for population modeling that can be widely accepted by these various communities. These standards include the language used for encoding a model, the approach for representing a model and the methods for using it:

- **The approach.** Our approach consists in seeing a model as hierarchical, represented by a joint probability distribution. This joint distribution can be decomposed into a product of conditional distributions, each associated with a submodel (model for observations, individual parameters, etc.). Tasks required of the modeler are thus related to these probability distributions.
- **The methods.** Many tests have shown that algorithms implemented in MONOLIX are the most reliable, all the while being extremely fast. In fact, these algorithms are precisely described and published in well known statistical journals. In particular, the SAEM algorithm, used for calculating the maximum likelihood estimation of population parameters, has shown its worth in numerous situations. Its mathematical convergence has also been proven under quite general hypotheses.
- **The language.** Mlxtran is used by MONOLIX and other modeling tools and is today by far the most advanced language for representing models. Initially developed for representing pharmacometric models, its syntax also allows it to easily code dynamical systems defined by a system of ODEs, and statistical models involving continuous, discrete and survival variables. This flexibility is a true advantage both for numerical modelers and statisticians.

3. Research Program

3.1. Scientific positioning

"Interfaces" is the defining characteristic of XPOP:

The interface between statistics, probability and numerical methods. Mathematical modelling of complex biological phenomena require to combine numerical, stochastic and statistical approaches. The CMAP is therefore the right place to be for positioning the team at the interface between several mathematical disciplines.

The interface between mathematics and the life sciences. The goal of XPOP is to bring the right answers to the right questions. These answers are mathematical tools (statistics, numerical methods, etc.), whereas the questions come from the life sciences (pharmacology, medicine, biology, etc.). This is why the point of XPOP is not to take part in mathematical projects only, but also pluridisciplinary ones.

The interface between mathematics and software development. The development of new methods is the main activity of XPOP. However, new methods are only useful if they end up being implemented in a software tool. On one hand, a strong partnership with Lixoft (the spin-off company who continue developing MONOLIX) allows us to maintaining this positioning. On the other hand, several members of the team are very active in the R community and develop widely used packages.

3.2. The mixed-effects models

Mixed-effects models are statistical models with both fixed effects and random effects. They are well-adapted to situations where repeated measurements are made on the same individual/statistical unit.

Consider first a single subject i of the population. Let $y_i = (y_{ij}, 1 \leq j \leq n_i)$ be the vector of observations for this subject. The model that describes the observations y_i is assumed to be a parametric probabilistic model: let $p_Y(y_i; \psi_i)$ be the probability distribution of y_i, where ψ_i is a vector of parameters.

In a population framework, the vector of parameters ψ_i is assumed to be drawn from a population distribution $p_\Psi(\psi_i; \theta)$ where θ is a vector of population parameters.

Then, the probabilistic model is the joint probability distribution

$$p(y_i, \psi_i; \theta) = p_Y(y_i|\psi_i) p_\Psi(\psi_i; \theta) \tag{8}$$

To define a model thus consists in defining precisely these two terms.

In most applications, the observed data y_i are continuous longitudinal data. We then assume the following representation for y_i:

$$y_{ij} = f(t_{ij}, \psi_i) + g(t_{ij}, \psi_i)\varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i. \tag{9}$$

Here, y_{ij} is the observation obtained from subject i at time t_{ij}. The residual errors (ε_{ij}) are assumed to be standardized random variables (mean zero and variance 1). The residual error model is represented by function g in model (2).

Function f is usually the solution to a system of ordinary differential equations (pharmacokinetic/pharmacodynamic models, etc.) or a system of partial differential equations (tumor growth, respiratory system, etc.). This component is a fundamental component of the model since it defines the prediction of the observed kinetics for a given set of parameters.

The vector of individual parameters ψ_i is usually function of a vector of population parameters ψ_{pop}, a vector of random effects $\eta_i \sim \mathcal{N}(0, \Omega)$, a vector of individual covariates c_i (weight, age, gender, ...) and some fixed effects β.

The joint model of y and ψ depends then on a vector of parameters $\theta = (\psi_{\text{pop}}, \beta, \Omega)$.

3.3. Computational Statistical Methods

Central to modern statistics is the use of probabilistic models. To relate these models to data requires the ability to calculate the probability of the observed data: the likelihood function, which is central to most statistical methods and provides a principled framework to handle uncertainty.

The emergence of computational statistics as a collection of powerful and general methodologies for carrying out likelihood-based inference made complex models with non-standard data accessible to likelihood, including hierarchical models, models with intricate latent structure, and missing data.

In particular, algorithms previously developed by POPIX for mixed effects models, and today implemented in several software tools (especially MONOLIX) are part of these methods:

- the adaptive Metropolis-Hastings algorithm allows one to sample from the conditional distribution of the individual parameters $p(\psi_i|y_i; c_i, \theta)$,
- the SAEM algorithm is used to maximize the observed likelihood $\mathcal{L}(\theta; y) = p(y; \theta)$,
- Importance Sampling Monte Carlo simulations provide an accurate estimation of the observed log-likelihood $\log(\mathcal{L}(\theta; y))$.

Computational statistics is an area which remains extremely active today. Recently, one can notice that the incentive for further improvements and innovation comes mainly from three broad directions: the high dimensional challenge, the quest for adaptive procedures that can eliminate the cumbersome process of tuning "by hand" the settings of the algorithms and the need for flexible theoretical support, arguably required by all recent developments as well as many of the traditional MCMC algorithms that are widely used in practice.

Working in these three directions is a clear objective for XPOP.

3.4. Markov Chain Monte Carlo algorithms

While these Monte Carlo algorithms have turned into standard tools over the past decade, they still face difficulties in handling less regular problems such as those involved in deriving inference for high-dimensional models. One of the main problems encountered when using MCMC in this challenging settings is that it is difficult to design a Markov chain that efficiently samples the state space of interest.

The Metropolis-adjusted Langevin algorithm (MALA) is a Markov chain Monte Carlo (MCMC) method for obtaining random samples from a probability distribution for which direct sampling is difficult. As the name suggests, MALA uses a combination of two mechanisms to generate the states of a random walk that has the target probability distribution as an invariant measure:

1. new states are proposed using Langevin dynamics, which use evaluations of the gradient of the target probability density function;
2. these proposals are accepted or rejected using the Metropolis-Hastings algorithm, which uses evaluations of the target probability density (but not its gradient).

Informally, the Langevin dynamics drives the random walk towards regions of high probability in the manner of a gradient flow, while the Metropolis-Hastings accept/reject mechanism improves the mixing and convergence properties of this random walk.

Several extensions of MALA have been proposed recently by several authors, including fMALA (fast MALA), AMALA (anisotropic MALA), MMALA (manifold MALA), position-dependent MALA (PMALA), ...

MALA and these extensions have demonstrated to represent very efficient alternative for sampling from high dimensional distributions. We therefore need to adapt these methods to general mixed effects models.

3.5. Parameter estimation

The Stochastic Approximation Expectation Maximization (SAEM) algorithm has shown to be extremely efficient for maximum likelihood estimation in incomplete data models, and particularly in mixed effects models for estimating the population parameters. However, there are several practical situations for which extensions of SAEM are still needed:

High dimensional model: a complex physiological model may have a large number of parameters (in the order of 100). Then several problems arise:

- when most of these parameters are associated with random effects, the MCMC algorithm should be able to sample, for each of the N individuals, parameters from a high dimensional distribution. Efficient MCMC methods for high dimensions are then required.
- Practical identifiability of the model is not ensured with a limited amount of data. In other words, we cannot expect to be able to properly estimate all the parameters of the model, including the fixed effects and the variance-covariance matrix of the random effects. Then, some random effects should be removed, assuming that some parameters do not vary in the population. It may also be necessary to fix the value of some parameters (using values from the literature for instance). The strategy to decide which parameters should be fixed and which random effects should be removed remains totally empirical. XPOP aims to develop a procedure that will help the modeller to take such decisions.

Large number of covariates: the covariate model aims to explain part of the inter-patient variability of some parameters. Classical methods for covariate model building are based on comparisons with respect to some criteria, usually derived from the likelihood (AIC, BIC), or some statistical test (Wald test, LRT, etc.). In other words, the modelling procedure requires two steps: first, all possible models are fitted using some estimation procedure (e.g. the SAEM algorithm) and the likelihood of each model is computed using a numerical integration procedure (e.g. Monte Carlo Importance Sampling); then, a model selection procedure chooses the "best" covariate model. Such a strategy is only possible with a reduced number of covariates, i.e., with a "small" number of models to fit and compare.

As an alternative, we are thinking about a Bayesian approach which consists of estimating simultaneously the covariate model and the parameters of the model in a single run. An (informative or uninformative) prior is defined for each model by defining a prior probability for each covariate to be included in the model. In other words, we extend the probabilistic model by introducing binary variables that indicate the presence or absence of each covariate in the model. Then, the model selection procedure consists of estimating and maximizing the conditional distribution of this sequence of binary variables. Furthermore, a probability can be associated to any of the possible covariate models.

This conditional distribution can be estimated using an MCMC procedure combined with the SAEM algorithm for estimating the population parameters of the model. In practice, such an approach can only deal with a limited number of covariates since the dimension of the probability space to explore increases exponentially with the number of covariates. Consequently, we would like to have methods able to find a small number of variables (from a large starting set) that influence certain parameters in populations of individuals. That means that, instead of estimating the conditional distribution of all the covariate models as described above, the algorithm should focus on the most likely ones.

Fixed parameters: it is quite frequent that some individual parameters of the model have no random component and are purely fixed effects. Then, the model may not belong to the exponential family anymore and the original version of SAEM cannot be used as it is. Several extensions exist:

- introduce random effects with decreasing variances for these parameters,
- introduce a prior distribution for these fixed effects,
- apply the stochastic approximation directly on the sequence of estimated parameters, instead of the sufficient statistics of the model.

None of these methods always work correctly. Furthermore, what are the pros and cons of these methods is not clear at all. Then, developing a robust methodology for such model is necessary.

Convergence toward the global maximum of the likelihood: convergence of SAEM can strongly depend on thie initial guess when the observed likelihood has several local maxima. A kind of simulated annealing version of SAEM was previously developed and implemented in MONOLIX. The method works quite well in most situations but there is no theoretical justification and choosing the settings of this algorithm (i.e. how the temperature decreases during the iterations) remains empirical. A precise analysis of the algorithm could be very useful to better understand why it "works" in practice and how to optimize it.

Convergence diagnostic: Convergence of SAEM was theoretically demonstrated under very general hypothesis. Such result is important but of little interest in practice at the time to use SAEM in a finite amount of time, i.e. in a finite number of iterations. Some qualitative and quantitative criteria should be defined in order to both optimize the settings of the algorithm, detect a poor convergence of SAEM and evaluate the quality of the results in order to avoid using them unwisely.

3.6. Model building

Defining an optimal strategy for model building is far from easy because a model is the assembled product of numerous components that need to been evaluated and perhaps improved: the structural model, residual error model, covariate model, covariance model, etc.

How to proceed so as to obtain the best possible combination of these components? There is no magic recipe but an effort will be made to provide some qualitative and quantitative criteria in order to help the modeller for building his model.

The strategy to take will mainly depend on the time we can dedicate to building the model and the time required for running it. For relatively simple models for which parameter estimation is fast, it is possible to fit many models and compare them. This can also be done if we have powerful computing facilities available (e.g., a cluster) allowing large numbers of simultaneous runs.

However, if we are working on a standard laptop or desktop computer, model building is a sequential process in which a new model is tested at each step. If the model is complex and requires significant computation time (e.g., when involving systems of ODEs), we are constrained to limit the number of models we can test in a reasonable time period. In this context, it also becomes important to carefully choose the tasks to run at each step.

3.7. Model evaluation

Diagnostic tools are recognized as an essential method for model assessment in the process of model building. Indeed, the modeler needs to confront "his" model with the experimental data before concluding that this model is able to reproduce the data and before using it for any purpose, such as prediction or simulation for instance.

The objective of a diagnostic tool is twofold: first we want to check if the assumptions made on the model are valid or not ; then, if some assumptions are rejected, we want to get some guidance on how to improve the model.

As is the usual case in statistics, it is not because this "final" model has not been rejected that it is necessarily the "true" one. All that we can say is that the experimental data does not allow us to reject it. It is merely one of perhaps many models that cannot be rejected.

Model diagnostic tools are for the most part graphical, i.e., visual; we "see" when something is not right between a chosen model and the data it is hypothesized to describe. These diagnostic plots are usually based on the empirical Bayes estimates (EBEs) of the individual parameters and EBEs of the random effects: scatterplots of individual parameters versus covariates to detect some possible relationship, scatterplots of pairs of random effects to detect some possible correlation between random effects, plot of the empirical distribution of the random effects (boxplot, histogram,...) to check if they are normally distributed, ...

The use of EBEs for diagnostic plots and statistical tests is efficient with rich data, i.e. when a significant amount of information is available in the data for recovering accurately all the individual parameters. On the contrary, tests and plots can be misleading when the estimates of the individual parameters are greatly shrunk.

We propose to develop new approaches for diagnosing mixed effects models in a general context and derive formal and unbiased statistical tests for testing separately each feature of the model.

3.8. Missing data

The ability to easily collect and gather a large amount of data from different sources can be seen as an opportunity to better understand many processes. It has already led to breakthroughs in several application areas. However, due to the wide heterogeneity of measurements and objectives, these large databases often exhibit an extraordinary high number of missing values. Hence, in addition to scientific questions, such data also present some important methodological and technical challenges for data analyst.

Missing values occur for a variety of reasons: machines that fail, survey participants who do not answer certain questions, destroyed or lost data, dead animals, damaged plants, etc. Missing values are problematic since most statistical methods can not be applied directly on a incomplete data. Many progress have been made to properly handle missing values. However, there are still many challenges that need to be addressed in the future, that are crucial for the users.

- State of arts methods often consider the case of continuous or categorical data whereas real data are very often mixed. The idea is to develop a multiple imputation method based on a specific principal component analysis (PCA) for mixed data. Indeed, PCA has been used with success to predict (impute) the missing values. A very appealing property is the ability of the method to handle very large matrices with large amount of missing entries.
- The asymptotic regime underlying modern data is not any more to consider that the sample size increases but that both number of observations and number of variables are very large. In practice first experiments showed that the coverage properties of confidence areas based on the classical methods to estimate variance with missing values varied widely. The asymptotic method and the bootstrap do well in low-noise setting, but can fail when the noise level gets high or when the number of variables is much greater than the number of rows. On the other hand, the jackknife has good coverage properties for large noisy examples but requires a minimum number of variables to be stable enough.
- Inference with missing values is usually performed under the assumption of "Missing at Random" (MAR) values which means that the probability that a value is missing may depend on the observed data but does not depend on the missing value itself. In real data and in particular in data coming from clinical studies, both "Missing Non at Random" (MNAR) and MAR values occur. Taking into account in a proper way both types of missing values is extremely challenging but is worth investigating since the applications are extremely broad.

It is important to stress that missing data models are part of the general incomplete data models addressed by XPOP. Indeed, models with latent variables (i.e. non observed variables such as random effects in a mixed effects model), models with censored data (e.g. data below some limit of quantification) or models with dropout mechanism (e.g. when a subject in a clinical trial fails to continue in the study) can be seen as missing data models.

4. Application Domains

4.1. Surface Enhanced Raman Spectroscopy

(joint project with HEGP, AP-HP, and Lip(Sys)2, Université Paris-Saclay)

The objective of this work is to evaluate the feasibility of an evolving technique, surface enhanced Raman spectroscopy (SERS) for the analysis of cytotoxic drug concentration. This technique using silver nanoparticles was applied for quantitative analysis of 5-fluorouracil, one of the most widely used molecules in oncology [8].

In view of the high spectral variability observed between the various repetitions of the experiment, and the observed nonlinear interaction between signal concentration and intensity, nonlinear regression methods that take these variabilities into account have been developed.

4.2. Management of severe trauma

(joint project with the Traumabase group, AP-HP)

Major trauma is defined as any injury that endangers the life or the functional integrity of a person. It has been shown that management of major trauma based on standardized and protocol based care improves prognosis of patients especially for the two main causes of death in major trauma i.e., hemorrhage and traumatic brain injury.

However, evidence shows that patient management even in mature trauma systems often exceeds acceptable time frames, and despite existing guidelines deviations from protocol-based care are often observed. These deviations lead to a high variability in care and are associated with bad outcome such as inadequate hemorrhage control or delayed transfusion. Two main factors explain these observations. First, decision-making in trauma care is particularly demanding, because it requires rapid and complex decisions under time pressure in a very dynamic and multi-player environment characterized by high levels of uncertainty and stress. Second, being a complex and multiplayer process, trauma care is affected by fragmentation. Fragmentation is often the result of loss or deformation of information.

This disruptive influence prevents providers to engage with each other and commit to the care process.In order to respond to this challenge, our program has set the ambitious goal to develop a trauma decision support tool, the TraumaMatrix. The program aims to provide an integrative decision support and information management solution to clinicians for the first 24 hours of major trauma management. This program is divided into three steps.

Based on a detailed and high quality trauma database, Step 1 consists in developing the mathematical tools and models to predict trauma specific outcomes and decisions. This step raises considerable scientific and methodological challenges.

Step 2 will use these methods to apply them to develop in close cooperation with trauma care experts the decision support tool and develop a user friendly and ergonomic interface to be used by clinicians.

Step 3 will further develop the tool and interface and test in real-time its impact on clinician decision making and patient outcome.

4.3. Precision medicine and pharmacogenomics

(joint project with Dassault Systèmes)

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal–regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

4.4. Oncology

(joint project with the Biochemistry lab of Ecole Polytechnique and Institut Curie)

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients*.

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly though a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

4.5. Anesthesiology

(joint project with AP-HP Lariboisière and M3DISIM)

Two hundred million general anaesthesias are performed worldwide every year. Low blood pressure during anaesthesia is common and has been identified as a major factor in morbidity and mortality. These events require great reactivity in order to correct them as quickly as possible and impose constraints of reliability and reactivity to monitoring and treatment.

Recently, studies have demonstrated the usefulness of noradrelanine in preventing and treating intraoperative hypotension. The handling of this drug requires great vigilance with regard to the correct dosage. Currently, these drugs are administered manually by the healthcare staff in bolus and/or continuous infusion. This represents a heavy workload and suffers from a great deal of variability in order to find the right dosage for the desired effect on blood pressure.

The objective of this project is to automate the administration of noradrelanine with a closed-loop system that makes it possible to control the treatment in real time to an instantaneous blood pressure measurement.

4.6. Intracellular processes

(joint project with the InBio and IBIS inria teams and the MSC lab, UMR 7057)

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations [9].

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.
- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.
- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

4.7. Population pharmacometrics

(joint project with Lixoft)

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage [7], [21], [14].

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

4.8. Mass spectrometry

(joint project with the Molecular Chemistry Laboratory, LCM, of Ecole Polytechnique)

One of the main recent developments in analytical chemistry is the rapid democratization of high-resolution mass spectrometers. These instruments produce extremely complex mass spectra, which can include several hundred thousand ions when analyzing complex samples. The analysis of complex matrices (biological, agri-food, cosmetic, pharmaceutical, environmental, etc.) is precisely one of the major analytical challenges of this new century. Academic and industrial researchers are particularly interested in trying to quickly and effectively establish the chemical consequences of an event on a complex matrix. This may include, for example, searching for pesticide degradation products and metabolites in fruits and vegetables, photoproducts of active ingredients in a cosmetic emulsion exposed to UV rays or chlorination products of biocides in hospital effluents. The main difficulty of this type of analysis is based on the high spatial and temporal variability of the samples, which is in addition to the experimental uncertainties inherent in any measurement and requires a large number of samples and analyses to be carried out and computerized data processing (up to 16 million per mass spectrum).

A collaboration between XPOP and the Molecular Chemistry Laboratory (LCM) of the Ecole Polytechnique began in 2018. Our objective is to develop new methods for the statistical analysis of mass spectrometry data.

These methods are implemented in the SPIX software.

5. Highlights of the Year

5.1. Highlights of the Year

Version 2.0 of the SPIX software was available in June 2019.

5.1.1. Awards

Geneviève Robin was awarded: “Prix L’Oréal-UNESCO Pour les Femmes et la Science (Jeunes Talents France 2019)”

6. New Software and Platforms

6.1. mlxR

KEYWORDS: Simulation - Data visualization - Clinical trial simulator

FUNCTIONAL DESCRIPTION: The models are encoded using the model coding language 'Mlxtran', automatically converted into C++ codes, compiled on the fly and linked to R using the 'Rcpp' package. That allows one to implement very easily complex ODE-based models and complex statistical models, including mixed effects models, for continuous, count, categorical, and time-to-event data.

- Contact: Marc Lavielle
- URL: http://simulx.webpopix.org/

6.2. Rsmlx

R speaks Monolix

KEYWORDS: Data modeling - Nonlinear mixed effects models - Statistical modeling

FUNCTIONAL DESCRIPTION: Among other tasks, 'Rsmlx' provides a powerful tool for automatic PK model building, performs statistical tests for model assessment, bootstrap simulation and likelihood profiling for computing confidence intervals. 'Rsmlx' also proposes several automatic covariate search methods for mixed effects models.

- Partner: Lixoft
- Contact: Marc Lavielle
- URL: http://rsmlx.webpopix.org/

6.3. SPIX

KEYWORDS: Data modeling - Mass spectrometry - Chemistry

FUNCTIONAL DESCRIPTION: SPIX allows you to - automatically identify, on the basis of statistical approaches, small but significant differences in spectra measured under different conditions, - model the kinetics of entities that evolve over time

- Authors: Marc Lavielle and Yao Xu
- Partner: Laboratoire de Chimie Moléculaire - Ecole Polytechnique
- Contact: Marc Lavielle
- URL: http://spix.webpopix.org/

7. New Results

7.1. Modelling inheritance and variability of kinetic gene expression parameters in microbial cells

Modern experimental technologies enable monitoring of gene expression dynamics in individual cells and quantification of its variability in isogenic microbial populations. Among the sources of this variability is the randomness that affects inheritance of gene expression factors at cell division. Known parental relationships among individually observed cells provide invaluable information for the characterization of this extrinsic source of gene expression noise. Despite this fact, most existing methods to infer stochastic gene expression models from single-cell data dedicate little attention to the reconstruction of mother-daughter inheritance dynamics. Starting from a transcription and translation model of gene expression, we proposed a stochastic model for the evolution of gene expression dynamics in a population of dividing cells. Based on this model, we developed a method for the direct quantification of inheritance and variability of kinetic gene expression parameters from single-cell gene expression and lineage data. We demonstrated that our approach provides unbiased estimates of mother-daughter inheritance parameters, whereas indirect approaches using lineage information only in the post-processing of individual-cell parameters underestimate inheritance. Finally, we have shown on yeast osmotic shock response data that daughter cell parameters are largely determined by the mother, thus confirming the relevance of our method for the correct assessment of the onset of gene expression variability and the study of the transmission of regulatory factors [9].

7.2. Main effects and interactions in mixed and incomplete data frames

A mixed data frame (MDF) is a table collecting categorical, numerical and count observations. The use of MDF is widespread in statistics and the applications are numerous from abundance data in ecology to recommender systems. In many cases, an MDF exhibits simultaneously main effects, such as row, column or group effects and interactions, for which a low-rank model has often been suggested. Although the literature on low-rank approximations is very substantial, with few exceptions, existing methods do not allow to incorporate main effects and interactions while providing statistical guarantees. We proposed a new method that fills this gap [11], [3].

7.3. Quantification of gemcitabine intravenous drugs

This aim of this study was to assess the ability of Raman spectroscopy to quantify antineoplastic drugs directly in the finished product in plastic bags using a handled Raman spectrometer. Gemcitabine diluted in 0.9% sodium chloride was analyzed at various concentrations ranging from 1 to 20mg/mL directly through plastic bags using a handled 785nm Raman spectrometer.In accordance with EMA guidelines, quantitative models were developed to predict gemcitabine concentration in bag using partial least squares (PLS) regression. In order to evaluate the transposability of the developed Raman method and the routine method (flow injection analysis with UV detection), independent samples were analyzed using both techniques. The impact of the plastic bag was also evaluated by analysis samples through two different bags.The best model was obtained after standard normal variates preprocessing (SNV) for 15 latent variables. This model presented an excellent correlation between predicted and theoretical concentration values (R^2 of 0.9938 from the calibration set), a low limit of quantification (LLOQ) of 3.68mg/mL and acceptable repeatability and intermediate precision lower than the expected acceptance limit of 5% over the entire concentration range tested (except for the average concentration of 5.73mg/mL). For the 48 preparations higher than the LLOQ, the Bland-Altman approach showed the interchangeability of the wo methods with a difference bias of 2%. Moreover, no significant difference of predicted concentrations between the two containers tested ($p = 0.189$) was observed.Despite some limitations for low concentrations, this study clearly shows promising results for real-time monitoring of gemcitabine infusion preparations without removing samples. The non-invasive nature of this method should ensure the correct dose before administration to patients and with heightened safety for operators [8].

7.4. Low-rank model with covariates for count data analysis

Count data are collected in many scientific and engineering tasks including image processing, single-cell RNA sequencing and ecological studies. Such data sets often contain missing values, for example because some ecological sites cannot be reached in a certain year. In addition, in many instances, side information is also available, for example covariates about ecological sites or species. Low-rank methods are popular to denoise and impute count data, and benefit from a substantial theoretical background. Extensions accounting for covariates have been proposed, but to the best of our knowledge their theoretical and empirical properties have not been thoroughly studied, and few softwares are available for practitioners. We propose a complete methodology called LORI (Low-Rank Interaction), including a Poisson model, an algorithm, and automatic selection of the regularization parameter, to analyze count tables with covariates. We also derive an upper bound on the estimation error. We provide a simulation study with synthetic data, revealing empirically that LORI improves on state of the art methods in terms of estimation and imputation of the missing values. We illustrate how the method can be interpreted through visual displays with the analysis of a well-know plant abundance data set, and show that the LORI outputs are consistent with known results. Finally we demonstrate the relevance of the methodology by analyzing a waterbirds abundance table from the French national agency for wildlife and hunting management (ONCFS). The method is available in the R package lori on the Comprehensive Archive Network (CRAN), [10].

7.5. Imputation and low-rank estimation with Missing Non At Random data

Missing values challenge data analysis because many supervised and unsupervised learning methods cannot be applied directly to incomplete data. Matrix completion based on low-rank assumptions are very powerful solution for dealing with missing values. However, existing methods do not consider the case of informative missing values which are widely encountered in practice. We propose matrix completion methods to recover Missing Not At Random (MNAR) data. Our first contribution is to suggest a model-based estimation strategy by modelling the missing mechanism distribution. An EM algorithm is then implemented, involving a Fast Iterative Soft-Thresholding Algorithm (FISTA). Our second contribution is to suggest a computationally efficient surrogate estimation by implicitly taking into account the joint distribution of the data and the missing mechanism: the data matrix is concatenated with the mask coding for the missing values ; a low-rank structure for exponential family is assumed on this new matrix, in order to encode links between variables and missing mechanisms. The methodology that has the great advantage of handling different missing value mechanisms is robust to model specification errors, [22].

7.6. A mathematical model to predict BNP levels in hemodialysis patients

Clinical interpretation of B-Type Natriuretic Peptide (BNP) levels in hemodialysis patients (HD) for fluid management remains elusive. We conducted a retrospective observational monocentric study. We built a mathematical model to predict BNP levels, using multiple linear regressions, [12].

7.7. Analysis of the global convergence of (fast) incremental EM methods

The EM algorithm is one of the most popular algorithm for inference in latent data models. The original formulation of the EM algorithm does not scale to large data set, because the whole data set is required at each iteration of the algorithm. To alleviate this problem, Neal and Hinton (1998) have proposed an incremental version of the EM (iEM) in which at each iteration the conditional expectation of the latent data (E-step) is updated only for a mini-batch of observations. Another approach has been proposed by Cappé and Moulines (2009) in which the E-step is replaced by a stochastic approximation step, closely related to stochastic gradient. In this study, we analyzed incremental and stochastic version of the EM algorithm in a common unifying framework. We also introduced a new version incremental version, inspired by the SAGA algorithm by Defazio et al. (2014). We established non-asymptotic convergence bounds for global convergence, [15].

7.8. Efficient Metropolis-Hastings sampling for nonlinear mixed effects models

The ability to generate samples of the random effects from their conditional distributions is fundamental for inference in mixed effects models. Random walk Metropolis is widely used to conduct such sampling, but such a method can converge slowly for high dimension problems, or when the joint structure of the distributions to sample is complex. We proposed a Metropolis-Hastings (MH) algorithm based on a multidimensional Gaussian proposal that takes into account the joint conditional distribution of the random effects and does not require any tuning, in contrast with more sophisticated samplers such as the Metropolis Adjusted Langevin Algorithm or the No-U-Turn Sampler that involve costly tuning runs or intensive computation. Indeed, this distribution is automatically obtained thanks to a Laplace approximation of the original model. We have shown that such approximation is equivalent to linearizing the model in the case of continuous data, [14], [2].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Contract with Dassault Systèmes

Contract with Lixoft

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

Mixed-Effects Models of Intracellular Processes: Methods, Tools and Applications (MEMIP)

Coordinator: Gregory Batt (InBio Inria team)

Other partners: InBio and IBIS Inria teams, Laboratoire Matière et Systèmes Complexes (UMR 7057; CNRS and Paris Diderot Univ.)

9.1.2. Institut National du Cancer (INCa)

Targeting Rac-dependent actin polymerization in cutaneous melanoma - Institut National du Cancer

Coordinator: Alexis Gautreau (Ecole Polytechnique)

Other partners: Laboratoire de Biochimie (Polytechnique), Institut Curie, INSERM.

9.2. International Initiatives

9.2.1. International Initiatives

SaSMoTiDep

Title: Statistical and Stochastic modeling for time-dependent data

International Partners (Institution - Laboratory - Researcher):

Universidad de Valparaiso (Chile) - Centro de Investigación y Modelamiento de Fenómenos Aleatorios Valparaíso (CIMFAV) - Cristian Meza Becerra

Universidad Nacional de Colombia (Colombia) - Department of Statistics - Viswanathan Arunachalam

Duration: 01/01/2018 - 31/12/2019

Start year: 2018

See also: https://sasmotidep.uv.cl

In many applications, multiple measurements are made on one or several experimental units over a period of time. Such data could be called time-dependent data. From a statistical point of view, if we consider only one experimental unit, we can use a time series analysis. In the other hand, if we consider experimental designs (or observational studies) for several experimental units (or subjects) where each subject is measured at several points in time, we can use the term longitudinal data. In this project, we propose to study several statistical and stochastic models for repeated measures using parametric and non-parametric approaches. In particular, we will study the inference in complex mixed effects models, we will propose novel segmentation models for multiple series, non-parametric methods in dependent models and stochastic models. We will apply these methods to real data from several fields as biometrics, reliability, population dynamics and finance.

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Ricardo Rios, Universidad Central de Venezuela, Caracas: September 2019.

Cristian Meza, Universidad de Valparaiso, Chile, September 2019.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

- ALT 2019
- CLAPEM 2019, Merida, Mexico
- useR!2019

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Stochastic Processes and their Applications
- Journal of Statistical Planning and Inference
- Journal of Computational and Graphical Statistics

10.1.3. Invited Talks

- International Federation of Classification Societies, Thessaloniki, Greece, August 2019.
- useR!2019, Toulouse, France, July 2019.
- R in Official Statistics, Bucharest, Romania, May 2019.
- French Statistical Society, France, Nancy, June 2019.
- Statlearn 2019, Grenoble, France.
- The Big Sick Conference, Zermatt, Switzerland, February, 2019.
- REDIF 2019, La Habana, Cuba, October 2019.
- 4th Worshop on Virus Dynamics, Paris, October 2019.
- ModelBio 2019, Salamnca, Spain, November 2019.

10.1.4. Leadership within the Scientific Community

- Eric Moulines is in charge of the academic supervision of the International Laboratory of Stochastic Algorithms and High-dimensional inference, National Research University, Higher School of Economics, funded by the Russian Academic Excellence Project.
- Eric Moulines is associate researcher of the Alan Turing Institute
- Eric Moulines is elected member of the French Académie des Sciences.
- Julie Josse is visiting researcher at Google France

10.1.5. Scientific Expertise

- Marc Lavielle is member of the Scientific Committee of the High Council for Biotechnologies.
- Marc Lavielle is member of the evaluation committee of the International Center for Mathematics (CIMAT), Guanajuato, Mexico.
- Eric Moulines is member of the award committee of foundation "Charles Defforey".

10.1.6. Research administration

- Marc Lavielle is member of the Scientific Programming Committee (CPS) of the Institute Henri Poincaré (IHP).
- Eric Moulines is a board member of the Institut de Convergence DataIA.
- Julie Josse is elected member of the R foundation.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : Julie Josse, Statistics with R, 48, M2, X-HEC

Master : Eric Moulines, Regression models, 36, M2, X-HEC

Engineering School : Eric Moulines, Statistics, 36, 2A, X

Engineering School : Eric Moulines, Markov Chains, 36, 3A, X

Engineering School : Erwan Le Pennec, Statistics, 36, 2A, X

Engineering School : Erwan Le Pennec, Statistical Learning, 36, 3A, X

Engineering School : Marc Lavielle, Statistics in Action, 48, 3A, X

10.2.2. Supervision

PhD defended : Nicolas Brosse, June 2019, Eric Moulines

PhD defended : Geneviève Robin, June 2019, Julie Josse and Eric Moulines

PhD defended : Belhal Karimi, September 2019, Marc Lavielle and Eric Moulines

PhD in progress : Marine Zulian, October 2016, Marc Lavielle

PhD in progress : Wei Jiang , October 2017, Julie Josse and Marc Lavielle

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Eric Moulines was member of the Evaluation Committee of Inria.

10.3.2. Creation of media or tools for science outreach

Marc Lavielle developed and maintains the learning platform Statistics in Action. The purpose of this online learning platform is to show how statistics (and biostatistics) may be efficiently used in practice using R. It is specifically geared towards teaching statistical modelling concepts and applications for self-study. Indeed, most of the available teaching material tends to be quite "static" while statistical modelling is very much subject to "learning by doing".

Julie Josse (with Nicholas Tierney and Nathalie Vialaneix) developed and maintains the website R-miss-tastic, A resource website on missing data.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] N. BROSSE.*Around the Langevin Monte Carlo algorithm : extensions and applications*, Université Paris Saclay, June 2019, https://hal.inria.fr/tel-02430579

[2] B. KARIMI.*Non-Convex Optimization for Latent Data Models : Algorithms, Analysis and Applications*, Université Paris-Saclay, September 2019, https://tel.archives-ouvertes.fr/tel-02319140

[3] G. ROBIN.*Low-rank methods for heterogeneous and multi-source data*, Université Paris-Saclay, June 2019, https://tel.archives-ouvertes.fr/tel-02168204

Articles in International Peer-Reviewed Journal

[4] A. HAVET, M. LERASLE, É. MOULINES.*Density estimation for RWRE*, in "Mathematical Methods of Statistics", March 2019, https://arxiv.org/abs/1806.05839 [*DOI :* 10.3103/S1066530719010022], https://hal.archives-ouvertes.fr/hal-01815990

[5] F. HUSSON, J. JOSSE, B. NARASIMHAN, G. ROBIN.*Imputation of mixed data with multilevel singular value decomposition*, in "Journal of Computational and Graphical Statistics", 2019, https://arxiv.org/abs/1804.11087 , forthcoming [*DOI :* 10.1080/10618600.2019.1585261], https://hal.archives-ouvertes.fr/hal-01781291

[6] W. JIANG, J. JOSSE, M. LAVIELLE.*Logistic Regression with Missing Covariates – Parameter Estimation, Model Selection and Prediction*, in "Computational Statistics and Data Analysis", December 2019, 106907, https://arxiv.org/abs/1805.04602 , forthcoming [*DOI :* 10.1016/J.CSDA.2019.106907], https://hal.archives-ouvertes.fr/hal-01958835

[7] B. KARIMI, M. LAVIELLE, É. MOULINES.*f-SAEM: A fast Stochastic Approximation of the EM algorithm for nonlinear mixed effects models*, in "Computational Statistics and Data Analysis", July 2019, forthcoming [*DOI :* 10.1016/J.CSDA.2019.07.001], https://hal.inria.fr/hal-01958248

[8] L. LÊ, M. BERGE, A. TFAYLI, A. BAILLET-GUFFROY, P. PROGNON, A. DOWEK, E. CAUDRON.*Quantification of gemcitabine intravenous drugs by direct measurement in chemotherapy plastic bags using a handheld Raman spectrometer*, in "Talanta", May 2019, vol. 196, p. 376-380 [*DOI :* 10.1016/J.TALANTA.2018.11.062], https://hal.archives-ouvertes.fr/hal-01970020

[9] A. MARGUET, M. LAVIELLE, E. CINQUEMANI.*Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data*, in "Bioinformatics", 2019, vol. 35, n^o 14, p. i586-i595 [*DOI :* 10.1093/BIOINFORMATICS/BTZ378], https://hal.archives-ouvertes.fr/hal-02317115

[10] G. ROBIN, J. JOSSE, É. MOULINES, S. SARDY.*Low-rank model with covariates for count data analysis*, in "Journal of Multivariate Analysis", April 2019, vol. 173, https://arxiv.org/abs/1703.02296 , https://hal.archives-ouvertes.fr/hal-01482773

[11] G. ROBIN, O. KLOPP, J. JOSSE, É. MOULINES, R. TIBSHIRANI.*Main effects and interactions in mixed and incomplete data frames*, in "Journal of the American Statistical Association", June 2019 [*DOI :* 10.1080/01621459.2019.1623041], https://hal.archives-ouvertes.fr/hal-02423445

[12] M. TOUZOT, P. SERIS, C. MAHEAS, J. VANMASSENHOVE, A.-L. LANGLOIS, K. MOUBAKIR, S. LAPLANCHE, T. PETITCLERC, C. RIDEL, M. LAVIELLE.*A mathematical model to predict BNP levels in hemodialysis patients*, in "Nephrology", 2019 [*DOI :* 10.1111/NEP.13586], https://hal.archives-ouvertes.fr/hal-02127228

Invited Conferences

[13] B. KARIMI, B. MIASOJEDOW, É. MOULINES, H.-T. WAI.*Non-asymptotic Analysis of Biased Stochastic Approximation Scheme*, in "COLT 2019 - 32nd Annual Conference on Conference on Learning Theory", Phoenix, United States, 2019, p. 1 - 33, https://hal.inria.fr/hal-02127750

International Conferences with Proceedings

[14] B. KARIMI, M. LAVIELLE.*Efficient Metropolis-Hastings sampling for nonlinear mixed effects models*, in "BAYSM 2018 - Bayesian Young Statisticians Meeting", Warwick, United Kingdom, Bayesian Statistics and New Generations - Proceedings of BAYSM, Springer, 2019, https://hal.inria.fr/hal-01958247

[15] B. KARIMI, H.-T. WAI, É. MOULINES, M. LAVIELLE.*On the Global Convergence of (Fast) Incremental Expectation Maximization Methods*, in "NeurIPS 2019 - 33th Annual Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, https://hal.inria.fr/hal-02334656

Conferences without Proceedings

[16] V. AUDIGIER, F. HUSSON, J. JOSSE, M. RESCHE-RIGON.*Imputation multiple pour données mixtes par analyse factorielle*, in "JdS2019 - 51es Journées de Statistique de la Société Française de Statistique", Vandœuvre-lès-Nancy, France, Société Française de Statistique, June 2019, https://hal-agrocampus-ouest.archives-ouvertes.fr/hal-02355840

[17] T. LEVENT, P. PREUX, E. LE PENNEC, J. BADOSA, G. HENRI, Y. BONNASSIEUX.*Energy Management for Microgrids: a Reinforcement Learning Approach*, in "ISGT-Europe 2019 - IEEE PES Innovative Smart Grid Technologies Europe", Bucharest, France, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ISGTEUROPE.2019.8905538], https://hal.archives-ouvertes.fr/hal-02382232

Other Publications

[18] F. CHOULY, J. LOUBANI, A. LOZINSKI, B. MÉJRI, K. MERITO, S. PASSOS, A. PINEDA.*Computing bitangents for transmission belts*, January 2020, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02429962

[19] W. JIANG, M. BOGDAN, J. JOSSE, B. MIASOJEDOW, V. ROCKOVA.*Adaptive Bayesian SLOPE—High-dimensional Model Selection with Missing Values*, January 2020, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02430600

[20] J. JOSSE, N. PROST, E. SCORNET, G. VAROQUAUX.*On the consistency of supervised learning with missing values*, March 2019, https://arxiv.org/abs/1902.06931 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02024202

[21] B. KARIMI, M. LAVIELLE, É. MOULINES.*On the Convergence Properties of the Mini-Batch EM and MCEM Algorithms*, October 2019, working paper or preprint, https://hal.inria.fr/hal-02334485

[22] A. SPORTISSE, C. BOYER, J. JOSSE.*Imputation and low-rank estimation with Missing Non At Random data*, January 2019, https://arxiv.org/abs/1812.11409 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01964720