RESEARCH CENTER
Saclay - Île-de-France

FIELD

Activity Report 2016

Section New Results

Edition: 2017-08-25

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE
1. COMETE Project-Team 4
2. DATASHAPE Team 8
3. DEDUCTEAM Team 15
4. GRACE Project-Team 17
5. MEXICO Project-Team 22
6. PARSIFAL Project-Team 26
7. SPECFUN Project-Team 32
8. TOCCATA Project-Team 35
APPLIED MATHEMATICS, COMPUTATION AND SIMULATION
9. COMMANDS Project-Team 39
10. DEFI Project-Team 42
11. DISCO Project-Team 50
12. GAMMA3 Project-Team 58
13. GECO Project-Team 69
14. POEMS Project-Team 73
15. SELECT Project-Team 81
16. TAO Project-Team 87
17. TROPICAL Team 93
DIGITAL HEALTH, BIOLOGY AND EARTH
18. AMIB Project-Team 101
19. GALEN Project-Team 106
20. LIFEWARE Project-Team 110
21. M3DISIM Project-Team 115
22. PARIETAL Project-Team 121
23. XPOP Team 130
NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING
24. INFINE Project-Team 132
PERCEPTION, COGNITION AND INTERACTION
25. AVIZ Project-Team 136
26. CEDAR Team 143
27. DAHU Project-Team 145
28. EX-SITU Team 146
29. ILDA Project-Team 153
30. SMIS Project-Team 158

COMETE Project-Team

7. New Results

7.1. Foundations of information hiding

Information hiding refers to the problem of protecting private information while performing certain tasks or interactions, and trying to avoid that an adversary can infer such information. This is one of the main areas of research in Comète; we are exploring several topics, described below.

7.1.1. Axioms for Information Leakage

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, we studied in [17] information leakage axiomatically, showing important dependencies among different axioms. We 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.

7.1.2. Up-To Techniques for Generalized Bisimulation Metrics

Bisimulation metrics allow us to compute distances between the behaviors of probabilistic systems. In [18] we presented enhancements of the proof method based on bisimulation metrics, by extending the theory of up-to techniques to (pre)metrics on discrete probabilistic concurrent processes.

Up-to techniques have proved to be a powerful proof method for showing that two systems are bisimilar, since they make it possible to build (and thereby check) smaller relations in bisimulation proofs. We defined soundness conditions for up-to techniques on metrics, and studied compatibility properties that allow us to safely compose up-to techniques with each other. As an example, we derived the soundness of the up-to-bisimilarity-metric-and-context technique.

The study was carried out for a generalized version of the bisimulation metrics, in which the Kantorovich lifting is parametrized with respect to a distance function. The standard bisimulation metrics, as well as metrics aimed at capturing multiplicative properties such as differential privacy, are specific instances of this general definition.

7.1.3. Compositional methods for information-hiding

Systems concerned with information hiding often use randomization to obfuscate the link between the observables and the information to be protected. The degree of protection provided by a system can be expressed in terms of the probability of error associated with the inference of the secret information. In [12] we considered a probabilistic process calculus to specify such systems, and we studied how the operators affect the probability of error. In particular, we characterized constructs that have the property of not decreasing the degree of protection, and that can therefore be considered safe in the modular construction of these systems. As a case study, we applied these techniques to the Dining Cryptographers, and we derive a generalization of Chaum's strong anonymity result.

7.1.4. Differential Privacy Models for Location-Based Services

In [13], we considered the adaptation of differential privacy to the context of location-based services (LBSs), which personalize the information provided to a user based on his current position. Assuming that the LBS provider is queried with a perturbed version of the position of the user instead of his exact one, we relied on differential privacy to quantify the level of indistinguishability (i.e., privacy) provided by this perturbation with respect to the user's position. In this setting, the adaptation of differential privacy can lead to various models depending on the precise form of indistinguishability required. We discussed the set of properties that

hold for these models in terms of privacy, utility and also implementation issues. More precisely, we first introduced and analyzed one of these models, the (D,eps)-location privacy, which is directly inspired from the standard differential privacy model. In this context, we described a general probabilistic model for obfuscation mechanisms for the locations whose output domain is the Euclidean space E^2. In this model, we characterized the satisfiability conditions of (D,eps)-location privacy for a particular mechanism and also measured its utility with respect to an arbitrary loss function. Afterwards, we presented and analyzed symmetric mechanisms in which all locations are perturbed in a unified manner through a noise function, focusing in particular on circular noise functions. We proved that, under certain assumptions, the circular functions are rich enough to provide the same privacy and utility levels as other more complex (i.e., non-circular) noise functions, while being easier to implement. Finally, we extended our results to a generalized notion for location privacy, called 'l-privacy' capturing both (D,eps)-location privacy and also the notion of geo-indistinguishability recently introduced by Andrès, Bordenabe, Chatzikokolakis and Palamidessi.

7.1.5. Practical Mechanisms for Location Privacy

The continuously increasing use of location-based services poses an important threat to the privacy of users. A natural defense is to employ an obfuscation mechanism, such as those providing geo-indistinguishability, a framework for obtaining formal privacy guarantees that has become popular in recent years.

Ideally, one would like to employ an optimal obfuscation mechanism, providing the best utility among those satisfying the required privacy level. In theory optimal mechanisms can be constructed via linear programming. In practice, however, this is only feasible for a radically small number of locations. As a consequence, all known applications of geo-indistinguishability simply use noise drawn from a planar Laplace distribution.

In [23] we studied methods for substantially improving the utility of location obfuscation, while having practical applicability as a central constraint. We provided such solutions for both infinite (continuous or discrete) as well as large but finite domains of locations, using a Bayesian remapping procedure as a key ingredient. We evaluated our techniques in two real world complete datasets, without any restriction on the evaluation area, and showed important utility improvements wrt the standard planar Laplace approach.

7.1.6. Preserving differential privacy under finite-precision semantics

The approximation introduced by finite-precision representation of continuous data can induce arbitrarily large information leaks even when the computation using exact semantics is secure. Such leakage can thus undermine design efforts aimed at protecting sensitive information. In [14] we focussed on differential privacy, an approach to privacy that emerged from the area of statistical databases and is now widely applied also in other domains. In this approach, privacy is protected by adding noise to the values correlated to the private data. The typical mechanisms used to achieve differential privacy have been proved correct in the ideal case in which computations are made using infinite-precision semantics. We analyzed the situation at the implementation level, where the semantics is necessarily limited by finite precision, i.e., the representation of real numbers and the operations on them are rounded according to some level of precision. We showed that in general there are violations of the differential privacy property, and we studied the conditions under which we can still guarantee a limited (but, arguably, acceptable) variant of the property, under only a minor degradation of the privacy level. Finally, we illustrated our results on two examples: the standard Laplacian mechanism commonly used in differential privacy, and a bivariate version of it recently introduced in the setting of privacy-aware geolocation.

7.1.7. Quantifying Leakage in the Presence of Unreliable Sources of Information

Belief and min-entropy leakage are two well-known approaches to quantify information flow in security systems. Both concepts stand as alternatives to the traditional approaches founded on Shannon entropy and mutual information , which were shown to provide inadequate security guarantees. In [16] we unified the two concepts in one model so as to cope with the frequent (potentially inaccurate, misleading or outdated) attackers' side information about individuals on social networks, online forums, blogs and other forms of online communication and information sharing. To this end we proposed a new metric based on min-entropy that takes into account the adversary's beliefs.

7.1.8. On the Compositionality of Quantitative Information Flow

In the min-entropy approach to quantitative information flow, the leakage is defined in terms of a minimization problem, which, in the case of large systems, can be computationally rather heavy. The same happens for the recently proposed generalization called g-vulnerability. In [25] we studied the case in which the channel associated to the system can be decomposed into simpler channels, which typically happens when the observables consist of several components. Our main contribution is the derivation of bounds on the g-leakage of the whole system in terms of the g-leakages of its components. We also considered the particular cases of min-entropy leakage and of parallel channels, generalizing and systematizing results from the literature. We demonstrated the effectiveness of our method and evaluate the precision of our bounds using examples.

7.2. Foundations of Concurrency

Distributed systems have changed substantially in the recent past with the advent of phenomena like social networks and cloud computing. In the previous incarnation of distributed computing the emphasis was on consistency, fault tolerance, resource management and related topics; these were all characterized by *interaction between processes*. Research proceeded along two lines: the algorithmic side which dominated the Principles Of Distributed Computing conferences and the more process algebraic approach epitomized by CONCUR where the emphasis was on developing compositional reasoning principles. What marks the new era of distributed systems is an emphasis on managing access to information to a much greater degree than before.

7.2.1. Belief, Knowledge, Lies and Other Utterances in an Algebra for Space and Extrusion

Spatial constraint systems are algebraic structures from concurrent constraint programming to specify spatial and epistemic behavior in multi-agent system. In [15], [11] we developed the theory of spatial constraint systems with operators to specify information and processes moving from a space to another. We investigated the properties of this new family of constraint systems and illustrated their applications. From a computational point of view the new operators provide for process/information extrusion, a central concept in formalisms for mobile communication. From an epistemic point of view extrusion corresponds to a notion we called utterance; a piece of information that an agent communicates to others but that may be inconsistent with the agent's beliefs. Utterances can then be used to express instances of epistemic notions such as hoaxes or intentional lies. Spatial constraint system can express the epistemic notion of belief by means of space functions that specify local information. We showed that spatial constraint can also express the epistemic notion of knowledge by means of a derived spatial operator that specifies global information. In [21] we reported on our progress using spatial constraint system as an abstract representation of modal and epistemic behaviour.

7.2.2. Deriving Inverse Operators for Modal Logic

In [20] we used spatial constraint systems to give an abstract characterization of the notion of normality in modal logic and to derive right inverse/reverse operators for modal languages. In particular, we identified the weakest condition for the existence of right inverses and showed that the abstract notion of normality corresponds to the preservation of finite suprema. We applied our results to existing modal languages such as the weakest normal modal logic, Hennessy-Milner logic, and linear-time temporal logic. We also discussed our results in the context of modal concepts such as bisimilarity and inconsistency invariance.

7.2.3. D-SPACES: Implementing Declarative Semantics for Spatially Structured Information

In [22] we introduced D-SPACES, an implementation of constraint systems with space and extrusion operators. D-SPACES is coded as a c++11 library providing implementations for constraint systems, space functions and extrusion functions. D-SPACES provides property-checking methods as well as an implementation of a specific type of constraint systems (boolean algebras). We illustrated the implementation with a small social network where users post their beliefs and utter their opinions.

7.2.4. Slicing Concurrent Constraint Programs

Concurrent Constraint Programming (CCP) is a declarative model for concurrency where agents interact by telling and asking constraints (pieces of information) in a shared store. Some previous works have developed (approximated) declarative debuggers for CCP languages. However, the task of debugging concurrent programs remains difficult. In [19] we defined a dynamic slicer for CCP and we showed it to be a useful companion tool for the existing debugging techniques. Our technique starts by considering a partial computation (a trace) that shows the presence of bugs. Often, the quantity of information in such a trace is overwhelming, and the user gets easily lost, since she cannot focus on the sources of the bugs. Our slicer allows for marking part of the state of the computation and assists the user to eliminate most of the redundant information in order to highlight the errors. We showed that this technique can be tailored to timed variants of CCP. We also developed a prototypical implementation freely available for making experiments.

DATASHAPE Team

7. New Results

7.1. Algorithmic aspects of topological and geometric data analysis

7.1.1. An Efficient Representation for Filtrations of Simplicial Complexes

Participant: Jean-Daniel Boissonnat.

In collaboration with Karthik C.S. (Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Israel)

A filtration over a simplicial complex K is an ordering of the simplices of K such that all prefixes in the ordering are subcomplexes of K. Filtrations are at the core of Persistent Homology, a major tool in Topological Data Analysis. In order to represent the filtration of a simplicial complex, the entire filtration can be appended to any data structure that explicitly stores all the simplices of the complex such as the Hasse diagram or the recently introduced Simplex Tree by Boissonnat and Maria [Algorithmica '14]. However, with the popularity of various computational methods that need to handle simplicial complexes, and with the rapidly increasing size of the complexes, the task of finding a compact data structure that can still support efficient queries is of great interest.

This direction has been recently pursued for the case of maintaining simplicial complexes. For instance, Boissonnat et al. [SoCG '15] considered storing the simplices that are maximal for the inclusion and Attali et al. [IJCGA '12] considered storing the simplices that block the expansion of the complex. Nevertheless, so far there has been no data structure that compactly stores the *filtration* of a simplicial complex, while also allowing the efficient implementation of basic operations on the complex.

In this work [22], we propose a new data structure called the Critical Simplex Diagram (CSD) which is a variant of our work on the Simplex Array List (SAL) introduced in [SoCG '15]. Our data structure allows to store in a compact way the filtration of a simplicial complex, and allows for the efficient implementation of a large range of basic operations. Moreover, we prove that our data structure is essentially optimal with respect to the requisite storage space. Next, we show that the CSD representation admits the following construction algorithms.

- A new *edge-deletion* algorithm for the fast construction of Flag complexes, which only depends on the number of critical simplices and the number of vertices.
- A new *matrix-parsing* algorithm to quickly construct the relaxed strong Delaunay complexes, depending only on the number of witnesses and the dimension of the complex.

7.1.2. Discretized Riemannian Delaunay triangulations

Participants: Mael Rouxel-Labbé, Mathijs Wintraecken, Jean-Daniel Boissonnat.

Anisotropic meshes are desirable for various applications, such as the numerical solving of partial differential equations and graphics. In [27], we introduce an algorithm to compute discrete approximations of Riemannian Voronoi diagrams on 2-manifolds. This is not straightforward because geodesics, shortest paths between points, and therefore distances cannot in general be computed exactly. Our implementation employs recent developments in the numerical computation of geodesic distances and is accelerated through the use of an underlying anisotropic graph structure. We give conditions that guarantee that our discrete Riemannian Voronoi diagram is combinatorially equivalent to the Riemannian Voronoi diagram and that its dual is an embedded triangulation, using both approximate geodesics and straight edges. Both the theoretical guarantees on the approximation of the Voronoi diagram and the implementation are new and provide a step towards the practical application of Riemannian Delaunay triangulations.

7.1.3. Efficient and Robust Persistent Homology for Measures

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with M. Buchet (Tohoku University), D. Sheehy (Univ. Connecticut).

A new paradigm for point cloud data analysis has emerged recently, where point clouds are no longer treated as mere compact sets but rather as empirical measures. A notion of distance to such measures has been defined and shown to be stable with respect to perturbations of the measure. This distance can easily be computed pointwise in the case of a point cloud, but its sublevel-sets, which carry the geometric information about the measure, remain hard to compute or approximate. This makes it challenging to adapt many powerful techniques based on the Euclidean distance to a point cloud to the more general setting of the distance to a measure on a metric space. We propose an efficient and reliable scheme to approximate the topological structure of the family of sublevel-sets of the distance to a measure. We obtain an algorithm for approximating the persistent homology of the distance to an empirical measure that works in arbitrary metric spaces. Precise quality and complexity guarantees are given with a discussion on the behavior of our approach in practice [17].

7.1.4. Shallow Packings in Geometry

Participants: Kunal Dutta, Arijit Ghosh.

A merged paper with Ezra, Esther (School of Mathematics, Georgia Institute of Technology, Atlanta, U.S.A.)

We refine the bound on the packing number, originally shown by Haussler, for shallow geometric set systems. Specifically, let V be a finite set system defined over an n-point set X; we view V as a set of indicator vectors over the n-dimensional unit cube. A $delta$-separated set of V is a subcollection W, such that the Hamming distance between each pair $u, v \in W$ is greater than δ, where $\delta > 0$ is an integer parameter. The δ-packing number is then defined as the cardinality of the largest δ-separated subcollection of V. Haussler showed an asymptotically tight bound of $\Theta((n/delta)^d)$ on the δ-packing number if V has VC-dimension (or primal shatter dimension) d. We refine this bound for the scenario where, for any subset, $X' \subset X$ of size $m \leq n$ and for any parameter $1 \leq k \leq m$, the number of vectors of length at most k in the restriction of V to X' is only $O(m^{d_1} k^{d-d_1})$, for a fixed integer $d > 0$ and a real parameter $1 \leq d_1 \leq d$ (this generalizes the standard notion of bounded primal shatter dimension when $d_1 = d$). In this case when V is "k-shallow" (all vector lengths are at most k), we show that its δ-packing number is $O(n^{d_1} k^{d-d_1}/\delta^d)$, matching Haussler's bound for the special cases where $d_1 = d$ or $k = n$. We present two proofs, the first is an extension of Haussler's approach, and the second extends the proof of Chazelle, originally presented as a simplification for Haussler's proof. [21]

- A new *tight upper bound* for shallow-packings in δ-separated set systems of bounded primal shatter dimension.

7.1.5. On Subgraphs of Bounded Degeneracy in Hypergraphs

Participants: Kunal Dutta, Arijit Ghosh.

A k-uniform hypergraph has degeneracy bounded by d if every induced subgraph has a vertex of degree at most d. Given a k-uniform hypergraph $H = (V(H), E(H))$, we show there exists an induced subgraph of size at least

$$\sum_{v \in V(H)} \min 1, ck \left(\frac{d+1}{d_H(v)+1} \right)^{1/(k-1)},$$

where $c_k = 2^{-(1+\frac{1}{k-1})}\left(1-\frac{1}{k}\right)$ and $d_H(v)$ denotes the degree of vertex v in the hypergraph H. This extends and generalizes a result of Alon-Kahn-Seymour (Graphs and Combinatorics, 1987) for graphs, as well as a result of Dutta-Mubayi-Subramanian (SIAM Journal on Discrete Mathematics, 2012) for linear hypergraphs, to general k-uniform hypergraphs. We also generalize the results of Srinivasan and Shachnai (SIAM Journal on Discrete Mathematics, 2004) from independent sets (0-degenerate subgraphs) to d-degenerate subgraphs. We further give a simple non-probabilistic proof of the Dutta-Mubayi-Subramanian bound for linear k-uniform hypergraphs, which extends the Alon-Kahn-Seymour proof technique to hypergraphs. Our proof combines the random permutation technique of Bopanna-Caro-Wei (see e.g. The Probabilistic Method, N. Alon and J. H. Spencer; Dutta-Mubayi-Subramanian) and also Beame-Luby (SODA, 1990) together with a new local density argument which may be of independent interest. We also provide some applications in discrete geometry, and address some natural algorithmic questions. [28]

- A new algorithmic *lower bound* for largest d-degenerate subgraphs in k-uniform hypergraphs.

7.1.6. A Simple Proof of Optimal Epsilon Nets

Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Nabil Mustafa (Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

Showing the existence of ε-nets of small size has been the subject of investigation for almost 30 years, starting from the initial breakthrough of Haussler and Welzl (1987). Following a long line of successive improvements, recent results have settled the question of the size of the smallest ε-nets for set systems as a function of their so-called shallow-cell complexity.

In this paper we give a short proof of this theorem in the space of a few elementary paragraphs, showing that it follows by combining the ε-net bound of Haussler and Welzl (1987) with a variant of Haussler's packing lemma (1991).

This implies all known cases of results on unweighted ε-nets studied for the past 30 years, starting from the result of Matoušek, Seidel and Welzl (1990) to that of Clarkson and Varadajan (2007) to that of Varadarajan (2010) and Chan, Grant, Könemann and Sharpe (2012) for the unweighted case, as well as the technical and intricate paper of Aronov, Ezra and Sharir (2010). [40]

- A new *unified proof* for all known bounds on unweighted ε-nets studied in the last 30 years.

7.1.7. Combinatorics of Set Systems with Small Shallow Cell Complexity: Optimal Bounds via Packings

Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Bruno Jartoux and Nabil Mustafa (Université Paris-Est Marne-la-Vallée, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

The packing lemma of Haussler states that given a set system (X, R) with bounded VC dimension, if every pair of sets in R are 'far apart' (i.e., have large symmetric difference), then R cannot contain too many sets. This has turned out to be the technical foundation for many results in geometric discrepancy using the entropy method as well as recent work on set systems with bounded VC dimension. Recently it was generalized to the shallow packing lemma [Dutta-Ezra-Ghosh SoCG 2015, Mustafa DCG 2016], applying to set systems as a function of their shallow cell complexity. In this paper we present several new results and applications related to packings:

1. an optimal lower bound for shallow packings, thus settling the open question in Ezra (SODA 2014) and Dutta et al. (SoCG 2015),
2. improved bounds on Mnets, providing a combinatorial analogue to Macbeath regions in convex geometry (Annals of Mathematics, 1952),
3. simplifying and generalizing the main technical tool in Fox et al. (J. of the EMS, 2016).

Besides using the packing lemma and a combinatorial construction, our proofs combine tools from polynomial partitioning and the probabilistic method. [37]

- A new *optimal lower bound* for shallow packings.
- New *improved bounds* for M-nets - combinatorial analogs of Macbeath regions in convex geometry.

7.1.8. A new asymmetric correlation inequality for Gaussian measure

Participants: Kunal Dutta, Arijit Ghosh.

In collaboration with Nabil Mustafa (Université Paris-Est Marne-la-Vallée, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France.)

The Khatri-Šidák lemma says that for any Gaussian measure μ over $\mathbb{R}^n$, given a convex set K and a slab L, both symmetric about the origin, one has $\mu(K \cap L) \geq \mu(K)\mu(L)$. We state and prove a new asymmetric version of the Khatri-Šidák lemma when K is a symmetric convex body and L is a slab (not necessarily symmetric about the barycenter of K). Our result also extends that of Szarek and Werner (1999), in a special case.

- A new *asymmetric* inequality for gaussian measure. [38].

7.2. Statistical aspects of topological and geometric data analysis

7.2.1. Stability and Minimax Optimality of Tangential Delaunay Complexes for Manifold Reconstruction

Participant: Eddie Aamari.

In collaboration with C. Levrard (Univ. Paris Diderot).

we consider the problem of optimality in manifold reconstruction. A random sample $\mathbb{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^D$ composed of points lying on a d-dimensional submanifold M, with or without outliers drawn in the ambient space, is observed. Based on the tangential Delaunay complex, we construct an estimator$\widehat{M}$ that is ambient isotopic and Hausdorff-close to M with high probability. $\widehat{M}$ is built from existing algorithms. In a model without outliers, we show that this estimator is asymptotically minimax optimal for the Hausdorff distance over a class of submanifolds with reach condition. Therefore, even with no a priori information on the tangent spaces of M, our estimator based on tangential Delaunay complexes is optimal. This shows that the optimal rate of convergence can be achieved through existing algorithms. A similar result is also derived in a model with outliers. A geometric interpolation result is derived, showing that the tangential Delaunay complex is stable with respect to noise and perturbations of the tangent spaces. In the process, a denoising procedure and a tangent space estimator both based on local principal component analysis (PCA) are studied [32].

7.2.2. Rates in the Central Limit Theorem and diffusion approximation via Stein's Method

Participant: Thomas Bonis.

We present a way to apply Stein's method in order to bound the Wasserstein distance between a, possibly discrete, measure and another measure assumed to be the invariant measure of a diffusion operator. We apply this construction to obtain convergence rates, in terms of p-Wasserstein distance for $p \geq 2$, in the Central Limit Theorem in dimension 1 under precise moment conditions. We also establish a similar result for the Wasserstein distance of order 2 in the multidimensional setting. In a second time, we study the convergence of stationary distributions of Markov chains in the context of diffusion approximation, with applications to density estimation from geometric random graphs and to sampling using the Langevin Monte Carlo algorithm [33].

7.2.3. Rates of Convergence for Robust Geometric Inference

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with P. Massart (Univ. Paris Sud et Inria Select team).

Distances to compact sets are widely used in the field of Topological Data Analysis for inferring geometric and topological features from point clouds. In this context, the distance to a probability measure (DTM) has been introduced by Chazal et al. as a robust alternative to the distance a compact set. In practice, the DTM can be estimated by its empirical counterpart, that is the distance to the empirical measure (DTEM). In this paper we give a tight control of the deviation of the DTEM. Our analysis relies on a local analysis of empirical processes. In particular, we show that the rate of convergence of the DTEM directly depends on the regularity at zero of a particular quantile function which contains some local information about the geometry of the support. This quantile function is the relevant quantity to describe precisely how difficult is a geometric inference problem. Several numerical experiments illustrate the convergence of the DTEM and also confirm that our bounds are tight [19].

7.2.4. Data driven estimation of Laplace-Beltrami operator

Participants: Frédéric Chazal, Bertrand Michel, Ilaria Giulini.

Approximations of Laplace-Beltrami operators on manifolds through graph Laplacians have become popular tools in data analysis and machine learning. These discretized operators usually depend on bandwidth parameters whose tuning remains a theoretical and practical problem. In this paper, we address this problem for the unnormalized graph Laplacian by establishing an oracle inequality that opens the door to a well-founded data-driven procedure for the bandwidth selection. Our approach relies on recent results by Lacour and Massart on the so-called Lepski's method [26].

7.3. Topological approach for multimodal data processing

7.3.1. Persistence-based Pooling for Shape Pose Recognition

Participants: Thomas Bonis, Frédéric Chazal, Steve Oudot, Maksim Ovsjanikov.

We propose a novel pooling approach for shape classification and recognition using the bag-of-words pipeline, based on topological persistence, a recent tool from Topological Data Analysis. Our technique extends the standard max-pooling, which summarizes the distribution of a visual feature with a single number, thereby losing any notion of spatiality. Instead, we propose to use topological persistence, and the derived persistence diagrams, to provide significantly more informative and spatially sensitive characterizations of the feature functions, which can lead to better recognition performance. Unfortunately, despite their conceptual appeal, persistence diagrams are difficult to handle, since they are not naturally represented as vectors in Euclidean space and even the standard metric, the bottleneck distance is not easy to compute. Furthermore, classical distances between diagrams, such as the bottleneck and Wasserstein distances, do not allow to build positive definite kernels that can be used for learning. To handle this issue, we provide a novel way to transform persistence diagrams into vectors, in which comparisons are trivial. Finally, we demonstrate the performance of our construction on the Non-Rigid 3D Human Models SHREC 2014 dataset, where we show that topological pooling can provide significant improvements over the standard pooling methods for the shape pose recognition within the bag-of-words pipeline [23].

7.3.2. Structure and Stability of the 1-Dimensional Mapper

Participants: Steve Oudot, Mathieu Carrière.

Given a continuous function $f : X \to \mathbb{R}$ and a cover $\mathcal{I}$ of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(\mathcal{I})$. Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of f, it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. We propose a theoretical framework that relates the structure of the Mapper to the one of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of (in-)stability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover $\mathcal{I}$ goes to zero [25].

7.3.3. Decomposition of exact pfd persistence bimodules

Participants: Steve Oudot, Jérémy Cochoy.

We characterize the class of persistence modules indexed over $\mathbb{R}^2$ that are decomposable into summands whose support have the shape of a *block*—i.e. a horizontal band, a vertical band, an upper-right quadrant, or a lower-left quadrant. Assuming the modules are *pointwise finite-dimensional* (pfd), we show that they are decomposable into block summands if and only if they satisfy a certain local property called *exactness*. Our proof follows the same scheme as the proof of decomposition 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 stability theory for zigzags and interlevel-sets persistence modules, in which block-decomposable bimodules play a key part. Our results allow us to drop some of the conditions under which that theory holds, in particular the Morse-type conditions [39].

7.4. Experimental research and software development

7.4.1. Topological Microstructure Analysis Using Persistence Landscapes

Participant: Paweł Dłotko.

In collaboration with T. Wanner (George Mason University).

Phase separation mechanisms can produce a variety of complicated and intricate microstructures, which often can be difficult to characterize in a quantitative way. In recent years, a number of novel topological metrics for microstructures have been proposed, which measure essential connectivity information and are based on techniques from algebraic topology. Such metrics are inherently computable using computational homology, provided the microstructures are discretized using a thresholding process. However, while in many cases the thresholding is straightforward, noise and measurement errors can lead to misleading metric values. In such situations, persistence landscapes have been proposed as a natural topology metric. Common to all of these approaches is the enormous data reduction, which passes from complicated patterns to discrete information. It is therefore natural to wonder what type of information is actually retained by the topology. In the present paper, we demonstrate that averaged persistence landscapes can be used to recover central system information in the Cahn-Hilliard theory of phase separation. More precisely, we show that topological information of evolving microstructures alone suffices to accurately detect both concentration information and the actual decomposition stage of a data snapshot. Considering that persistent homology only measures discrete connectivity information, regardless of the size of the topological features, these results indicate that the system parameters in a phase separation process affect the topology considerably more than anticipated. We believe that the methods discussed in this paper could provide a valuable tool for relating experimental data to model simulations [36].

7.4.2. Topological analysis of the connectome of digital reconstructions of neural microcircuits

Participant: Paweł Dłotko.

In collaboration with K. Hess, L. Ran, H. Markram, E. Muller, M. Nolte, M. Reimann, M. Scolamiero, K. Turner (Univ. of Aberden, EPFL, Brain and Mind Institute).

A first draft digital reconstruction and simulation of a microcircuit of neurons in the neocortex of a two-week-old rat was recently published. Since graph-theoretical methods may not be sufficient to understand the immense complexity of the network formed by the neurons and their connections, we explored whether application of methods from algebraic topology can provide a novel and useful perspective on the structural and functional organization of the microcircuit. Structural topological analysis revealed that directed graphs representing the connectivity between neurons are significantly different from random graphs and that there exist an enormous number of simplicial complexes of different dimensions representing all-to-all connections within different sets of neurons, the most extreme motif of neuronal clustering reported so far in the brain. Functional topological analysis based on data from simulations confirmed the interest of a new approach to

studying the relationship between the structure of the connectome and its emergent functions. In particular, functional responses to different stimuli can readily be distinguished by topological methods. This study represents the first algebraic topological analysis of connectomics data from neural microcircuits and shows promise for general applications in network science.

7.4.3. A persistence landscapes toolbox for topological statistics

Participant: Paweł Dłotko.

In collaboration with P. Bubenik (University of Florida).

Topological data analysis provides a multiscale description of the geometry and topology of quantitative data. The persistence landscape is a topological summary that can be easily combined with tools from statistics and machine learning. We give efficient algorithms for calculating persistence landscapes, their averages, and distances between such averages. We discuss an implementation of these algorithms and some related procedures. These are intended to facilitate the combination of statistics and machine learning with topological data analysis. We present an experiment showing that the low-dimensional persistence landscapes of points sampled from spheres (and boxes) of varying dimensions differ.

7.5. Miscellaneous

7.5.1. Monotone Simultaneous Paths Embeddings in $\mathbb{R}^d$

Participant: Marc Glisse.

In collaboration with O. Devillers and S. Lazard (Inria Nancy), David Bremner (University of New Brunswick, Canada), Giuseppe Liotta (University of Perugia, Italy), Tamara Mchedlidze (KIT, Germany), Sue Whitesides (University of Victoria, Canada), Stephen Wismath (University of Lethbridge, Canada).

We study[24] the following problem: Given k paths that share the same vertex set, is there a simultaneous geometric embedding of these paths such that each individual drawing is monotone in some direction? We prove that for any dimension $d \geq 2$, there is a set of $d + 1$ paths that does not admit a monotone simultaneous geometric embedding.

DEDUCTEAM Team

6. New Results

6.1. Dedukti

A. Assaf, G. Burel, R. Cauderlier, D. Delahaye, G. Dowek, C. Dubois, F. Gilbert, P. Halmagrand, O. Hermant, and R. Saillard, have finished writing a general presentation of the Dedukti system. This paper is submitted for publication.

Under the supervision of P. Halmagrand and G. Burel, D. Pham worked on the conversion of TSTP proof traces, as produced by automated theorem provers such as E, Zipperposition or Vampire, into Dedukti proofs. To that purpose, he modified Zenon modulo so that it reads TSTP files and tries to reprove the proof steps given by the trace.

R. Cauderlier defended his PhD thesis on the translation of programming languages to Dedukti and interoperability of proof systems [11]. He also presented his work on the use of Dedukti for rewriting-based proof transformation [15] and on the translation of FoCaLiZe in Dedukti [16]

6.2. Proof theory

G. Dowek and Y. Jiang have finished a paper on co-inductive and inductive complementation of inference systems. This paper is submitted for publication.

The paper of G. Dowek on the introduction of rules and derivations in a logic course has been published [24].

F. Gilbert has finished a paper on the automated constructivization of proofs, to appear in the proceedings of FOSSACS'17.

F. Thiré is working on the translation of the Fermat little theorem proof written in Matita to a proof written in HOL. A part of this work is developed in its internship report [25]. He is continuing this translation during his PhD thesis.

6.3. B Method

The B Method is a formal method mainly used in the railway industry to specify and develop safety-critical software. To guarantee the consistency of a B project, one decisive challenge is to show correct a large amount of proof obligations, which are mathematical formulas expressed in a classical set theory extended with a specific type system. To improve automated theorem proving in the B Method, Pierre Halmagrand proposes [17], [12] to use a first-order sequent calculus extended with a polymorphic type system, which is in particular the output proof-format of the tableau-based automated theorem prover Zenon. After stating some modifications of the B syntax and defining a sound elimination of comprehension sets, he proposes a translation of B formulas into a polymorphic first-order logic format. Then, he introduces the typed sequent calculus used by Zenon, and shows that Zenon proofs can be translated to proofs of the initial B formulas in the B proof system.

6.4. Termination

F. Blanqui revised his paper on "size-based termination of higher-order rewrite systems" submitted to the Journal of Functional Programming [23]. This paper is concerned with the termination, in Church' simply-typed λ-calculus, of the combination of β-reduction and arbitrary user-defined rewrite rules fired using matching modulo α-congruence only. Several authors have devised termination criteria for fixpoint-based function definitions using deduction rules for bounding the size of terms inhabiting inductively defined types, where the size of a term is (roughly speaking) the set-theoretical height of the tree representation of its normal form. In the present paper, we extend this approach to rewriting-based function definitions and more general notions of size.

G. Dowek has finished writing a paper on the notion of model and its application to termination proofs for the $\lambda\Pi$-calculus modulo theory. This paper is submitted for publication.

6.5. Confluence

In $\lambda\Pi$modulo, congruences are expressed by rewrite rules that must enjoy precise properties, notably confluence, strong normalization, and type preservation. A difficulty is that these properties depend on each other in calculi of dependent types. To break the circularity, confluence is usually proved separately on untyped terms. A another difficulty then arises : computation do not terminate on untyped terms. A result of van Oostrom allows to show confluence of non-terminating left-linear higher-order rules, provided their critical pairs are development closed. This result was used for the encodings of HOL, Matita, and Coq up to version 8.4. Encoding the most recent version of Coq requires rules for universes that are confluent on open terms, while confluence on ground terms sufficed before. The encoding we recently developed for this new version of Coq has higher-order rules which are not left-linear, use pattern matching modulo associativity, commutativity and identity, and whose (joinable) critical pairs are not development closed. We have therefore developed a new powerful result for proving confluence of that sort of rules provided non-linear variables can only be instantiated by first-order expressions [18], [19].

6.6. Physics and computation

The paper of G. Dowek and P. Arrighi Free fall and cellular automata has been published [13]. As a sequel of this paper, G. Dowek and P. Arrighi have written a short note [22].

A. Díaz-Caro and G. Dowek have developed a new typing system for quantum λ-calculus allowing to distinguish between pure states and superpositions.

Under the supervision of S. Martiel and P. Arrighi, C. Chouteau worked on a particular notion of covariance in the model of causal graph dynamics. Causal graph dynamics are graph transformations constrained by Physics-inspired symmetries. The particular object of study of this internship was a restriction of this model to physical transformations of discrete geometrical spaces.

GRACE Project-Team

7. New Results

7.1. Faster elliptic and hyperelliptic curve cryptography

B. Smith made several contributions to the development of faster arithmetic on elliptic curves and genus 2 Jacobians in 2016. In joint work with C. Costello and P.-N. Chung, he gave a new, efficient, uniform, and constant-time scalar multiplication algorithm for genus 2 Jacobians exploiting fast Kummer surface arithmetic and features of differential addition chains; this was presented at SAC 2016. The theory in this article was the basis of a highly competitive implementation of key exchange and signatures for microcontroller platforms, in joint work with J. Renes, P. Schwabe, and L. Batina, presented at CHES 2016.

7.2. Quantum factoring

Integer factorization via Shor's algorithm is a benchmark problem for general quantum computers, but surprisingly little work has been done on optimizing the algorithm for use as a serious factoring tool once large quantum computers are built (rather than as a proof of concept). In the meantime, given the limited size of contemporary quantum computers and the practical difficulties involved in building them, any optimizations to quantum factoring algorithms can lead to significant practical improvements. In a new interdisciplinary project with physicists F. Grosshans and T. Lawson, F. Morain and B. Smith have derived a simple new quantum factoring algorithm for cryptographic integers; its expected runtime is lower than Shor's factoring algorithm, and it should also be easier to implement in practice [22].

7.3. Advances in point counting

Determining the number of points on an elliptic curve, or more generally on the Jacobian of an algebraic curve, is a classic problem in algorithmic number theory that is now crucial for efficiently generating secure cryptographic parameters. Together with C. Scribot, F. Morain and B. Smith developed an improved version of the state-of-the-art SEA algorithm for certain families of elliptic curves with special endomorphisms; this was presented at ANTS-XII [10]. B. Smith also led a project group on special genus-2 point counting algorithms at the "Algebraic Geometry for Coding Theory and Cryptography" workshop at IPAM, UCLA, in 2016.

7.4. Cryptanalysis of code based cryptosystems by filtration attacks

The McEliece encryption scheme based on binary Goppa codes was one of the first public-key encryption schemes [31]. Its security rests on the difficulty of decoding an arbitrary code. The original proposal uses classical Goppa codes, and while it still remains unbroken, it requires a huge size of key. On the other hand, many derivative systems based on other families of algebraic codes have been subject to key recovery attacks. Up to now, key recovery attacks were based either on a variant of Sidelnikov and Shestakov's attack [32], where the first step involves the computation of minimum-weight codewords, or on the resolution of a system of polynomial equations using Gröbner bases.

In [26], A. Couvreur, P. Gaborit, V. Gauthier, A. Otmani and J.-P. Tillich introduced a new paradigm of attack called *filtration attacks*. The general principle decomposes in two steps:

1. **Distinguishing** the public code from a random one using the square code operation.
2. **Computing a filtration** of the public code using the distinguisher, and deriving from this filtration an efficient decoding algorithm for the public code.

This new style of attack allowed A. Couvreur, A. Otmani and J.-P. Tillich to break (in polynomial time) McEliece based on wild Goppa codes over quadratic extensions [3]. A detailed long version has been written and recently published [9]. A. Couvreur, Irene Márquez–Corbella, and R. Pellikaan broke McEliece based on algebraic geometry codes from curves of arbitrary genus [2], [27] by reconstructing optimal polynomial time decoding algorithms decoding up to the half minimum distance minus half the genus. This can be computed from the raw data of a generator matrix. In a recently submitted long version [21] the algorithm has been improved and permits to reconstruct a decoding algorithm up to the half minimum distance.

7.5. Quantum LDPC codes

Quantum codes are the analogous of error correcting codes for a quantum computer. A well known family of quantum codes are the CSS codes due to Calderbank, Shor and Steane can be represented by a pair of matrices (H_X, H_Z) such that $H_X H_Z^T = 0$. As in classical coding theory, if these matrices are sparse, then the code is said to be LDPC. An open problem in quantum coding theory is to get a family of quantum LDPC codes whose asymptotic minimum distance is in $\Omega(n^\alpha)$ for some $\alpha > 1/2$. No such family is known and actually, only few known families of quantum LDPC codes have a minimum distance tending to infinity.

In [24], Benjamin Audoux (I2M, Marseille) and A. Couvreur investigate a problem suggested by Bravyi and Hastings. They studied the behaviour of iterated tensor powers of CSS codes and prove in particular that such families always have a minimum distance tending to infinity. They propose also 3 families of LDPC codes whose minimum distance is in $\Omega(n^\beta)$ for all $\beta < 1/2$.

7.6. Discrete Logarithm computations in finite fields with the NFS algorithm

The best discrete logarithm record computations in prime fields and large characteristic finite fields are obtained with Number Field Sieve algorithm (NFS) at the moment. This algorithm is made of four steps:

1. polynomial selection;
2. relation collection (with a sieving technique);
3. linear algebra (computing the kernel of a huge matrix, of millions of rows and columns);
4. individual discrete logarithm computation.

The two more time consuming steps are the relation collection step and the linear algebra step. The polynomial selection is quite fast but is very important since it determines the complexity of the algorithm. Selecting better polynomials is a key to improve the overall running-time of the NFS algorithm.

A. Guillevic and F. Morain have written a chapter [18] on discrete logarithm computations for a book on pairings.

7.6.1. Breaking a MNT curve using DL computations

There is a reduction between an elliptic curve E defined over $\mathbf{F}_p$ and a finite extension of degree k (aka *embedding degree*) of the base field, using pairing computations. In brief, one can transport the discrete logarithm problem from E to $\mathbf{F}_{p^k}$. If k is relatively small, this yields a DLP much easier to solve than directly on E. To give some highlight on current easyness, A. Guillevic, F. Morain and E. Thomé (from CARAMBA EPC in LORIA) computed a discrete log on a curve of embedding degree 3 and cryptographic size. This clearly showed that curves with small embedding degrees are indeed weak. The article [14] was presented by A. Guillevic during the SAC 2016 conference in New Foundland.

7.7. Rank metric codes over infinite fields

Rank metric and Gabidulin codes over the rationals promise interesting applications to space-time coding. We have constructed optimal codes, similar to Gabidulin codes, in the case of infinite fields. We use algebraic extensions, and we have determined the condition on the considered extension to enable this construction. For example: we can design codes with complex coefficients, using number fields and Galois automorphisms. Then, in the rank metric setting, codewords can be seen as matrices. In this setting, a channel introduces

errors (a matrix of small rank r added to the codeword) and erasures (s_r rows and s_c columns of the matrix are erased). We have developed an algorithm (adapted from the Welch–Berlekamp algorithm) to recover the right codeword in the presence of an error of rank weight up to $r + s_c + s_r \leqslant d - 1$, where d is the minimal distance of the code. As opposed to the finite field case, we are confronted by coefficient size growth. We solve this problem by computing modulo prime ideals. Using these codes we can completely bypass intermediate constructions using finite fields, which were the stumbling-block in classic constructions.

We also have used this framework to build rank-metric codes over the field of rational functions, using algebraic function fields with cyclic Galois group (Kummer and Artin extensions). These codes can be seen as a generator of infinitely many convolutional codes.

7.8. Hash function cryptanalysis

Cryptographic hash functions are versatile primitives that are used in many cryptographic protocols. The security of a hash function h is usually evaluated through two main notions: its preimage resistance (given a target t, the difficulty of finding a message m s.t. $h(m) = t$) and its collision resistance (the difficulty of finding two messages m, m' s.t. $h(m) = h(m')$).

A popular hash function is the SHA-1 algorithm. Although theoretical collision attacks were found in 2005, it is still being used in some applications, for instance as the hash function in some TLS certificates. Hence cryptanalysis of SHA-1 is still a major topic in cryptography.

In 2015, we improved the state-of-the-art on SHA-1 analysis in two ways:

- T. Espitau, P.-A. Fouque and P. Karpman improved the previous preimage attacks on SHA-1, reaching up to 62 rounds (out of 80), up from 57. The corresponding paper was published at CRYPTO 2015.
- P. Karpman, T. Peyrin and M. Stevens developed collision attacks on the compression function of SHA-1 (i.e. freestart collisions). This exploits a model that is slightly more generous to the attacker in order to find explicit collisions on more rounds than what was previously possible. A first work resulted in freestart collisions for SHA-1 reduced to 76 steps; this attack takes less than a week to compute on a common GPU. The corresponding paper was published at CRYPTO 2015. This was later improved to attack the full compression function. Although the attack is more expensive it is still practical, taking less than two weeks on a 64 GPU cluster. The corresponding paper was accepted at EUROCRYPT 2016 [17].

7.9. Block cipher design and analysis

Block ciphers are one of the most basic cryptographic primitives, yet block cipher analysis is still a major research topic. In recent years, the community also shifted focus to the more general setting of *authenticated encryption*, where one specifies an (set of) algorithm(s) providing both encryption and authentication for messages of arbitrary length. A major current event in that direction is the CAESAR academic competition, which aims to select a portfolio of good algorithms.

In 2015, we helped to improve the state of the art in block cipher research in several ways:

- P. Karpman developed a compact 8-bit S-box with branch number three, which can be used as a basis to construct a lightweight block cipher particularly efficient on 8-bit microcontrollers [23].

In 2016, together with P.-A. Fouque, P. Kirchner and B. Minaud, P. Karpman designed a family of efficient provably incompressible symmetric primitives, which corresponds to a weak notion of white-box cryptography. The objective of such algorithms is that given an implementation of a certain target size, an adversary shouldn't be able to efficiently find a smaller implementation with comparable functionality. We introduced a security model that captures the behaviour of realistic adversaries and used this model to prove the security of a family of block cipher and a family of key generating functions. The corresponding paper was published at ASIACRYPT 2016 [13].

7.10. Weight distribution of Algebraic-Geometry codes

V. Ducet worked on the weight distribution of geometric codes following a method initiated by Duursma. More precisely he implemented his method in magma and was able to compute the weight distribution of the geometric codes coming from two optimal curves of genus 2 and 3 over the finite fields of size 16 and 9 respectively. The aim is to compute the weight distribution of the Hermitian code over the finite field of size 16, for which computational improvements of the implementation are necessary.

7.11. Update on the Chor-Rivest cryptosystem

The Chor-Rivest cryptosystem from the 90's was "broken" by Vaudenay. However, Vaudenay's attack applies only for the range of parameters originally proposed. The major recent breakthrough in discrete logarithm computations enable to redesign the system with a completly different range of paramaters, possibly thwarting Vaudenay's attack. D. Augot and C. Barbin tried to find a new attack against this discrete log and knapsack-based cryptosystem, using the Sidelnikov-Shestakov algorithm for recovering a Reed-Solomon code. Apparently, our new attack does not outperform S. Vaudenay's original attack, and it may be possible that the Chor-Rivest could be redesigned in a secure way.

7.12. Proofs or Retrievability

A Proof of Retrievability (PoR) is a cryptographic protocol which aims at ensuring a user that he can retrieve files he previously stored on a server. J. Lavauzelle and F. Levy-dit-Vehel studied a new approach for the construction of PoRs. The idea is to encode the file so that the user can check with low communication whether its file has been damaged. Such an encoding can be efficiently done with locally decodable and testable codes, and especially with the family of lifted codes introduced by Guo, Kopparty and Sudan [30]. In practice, PoRs thus defined achieve very efficient storage overhead and acceptable communication, compared to the existing litterature. This new construction [15] has been presented during the ISIT2016 conference in Barcelona.

7.13. Fast Encoding of Multiplicity Codes

N. Coxon has produced a fast implementation which demonstrates that the multiplicity codes from Kopparty, Saraf and Yehkanin are indeed practical for very large databases (when used in the Private Information Retrieval setting). For instance, we can encode a 10^8 bit longs message in two seconds on a regular laptop, and 10^9 in thirty seconds. We envisioned a scenario where DNA sequences are encoded using these multiplicity codes: 10^8 bits is the size of Drosophila melanogaster (flies), and 10^9 bits is the order of magnitude of the human genome.

7.14. Private Information Retrieval

Imagine the following scenario, in which a researcher wants to access many sustrings a DNA sequences, while maintaining the privacy of the request. The privacy or the secrecy of the database is not a concern here: for instance, this researcher wants to access many DNA subsequences of drosophila melanogaster, hosted on a remote data broker, and clearly the concern is not to protect the private life of flies. But the information leaked about the queries may endanger the novel aspect of the discovery the researcher is about to make, by revealing which DNA sequences he is studying.

Private Information Retrieval (PIR) schemes are designed to achieve this goal: a user queries a database T hosted on a remote server, and wants the i-th entry, i.e. $T[i]$. A cryptographic protocol is then run, and at the end of the protocol, the server must not know i, neither the $T[i]$ he answered, yet the user gets $T[i]$.

These PIR schemes can be achieved in an unconditionally secure way using the above Multiplicity codes, which N. Coxon made practical. In September, we explained this scenario and demoed our software at Nokia Bell Lab's Future X days a use case of Multiplicity codes for private access to DNA sequences.

7.15. Compact McEliece Keys from Algebraic-geometry codes

In 1978, McEliece [31], introduced a public key cryptosystem based on linear codes and suggested to use classical Goppa codes which belong to the family of alternant codes. This proposition remains secure but leads to very large public keys compared to other public-key cryptosystems. Many proposals have been made in order to reduce the key size, in particular quasi-cyclic alternant codes. Quasi-cyclic alternant codes refer to alternant codes admitting a generator matrix made of severals cyclic bloks. These alternant codes contains weakness because they have a non-trivial automorphism group. Thanks to this property we can build, from a quasi-cyclic alternant code, an alternant code with smaller parameters which has almost same private elements than the original code. Faugère, Otmani, Tillich, Perret and Portzamparc [29] showed this fact for alternant codes obtained by using supports $x \in \mathbb{F}_{q^m}^n$ globally stable by an affine map $\phi : z \mapsto az + b$, with $a, b \in \mathbb{F}_{q^m}^n$. E. Barelli has extended this proof to the non-affine case: for all codes obtained by using supports $x \in \mathbb{F}_{q^m}^n$ globally stable by a map $\phi : z \mapsto \frac{az+b}{cz+d}$, with $a, b, c, d \in \mathbb{F}_{q^m}^n$.

In order to suggest compact keys for the McEliece cryptosystem E. Barelli and A. Couvreur studied quasi-cyclic alternant gemeotric codes. Alternant geometric codes means a subfield subcode of an algebraic-geometry codes. To build these codes, we need curves with automorphisms. In particular, we studied Kummer cover of plane curves.

MEXICO Project-Team

7. New Results

7.1. Analyzing Timed Systems Using Tree Automata

Timed systems, such as timed automata, are usually analyzed using their operational semantics on timed words. The classical region abstraction for timed automata reduces them to (untimed) finite state automata with the same time-abstract properties, such as state reachability. In [10], we propose a new technique to analyze such timed systems using finite tree automata instead of finite word automata. The main idea is to consider timed behaviors as graphs with matching edges capturing timing constraints. Such graphs can be interpreted in trees opening the way to tree automata based techniques which are more powerful than analysis based on word automata. The technique is quite general and applies to many timed systems. In this paper, as an example, we develop the technique on timed pushdown systems, which have recently received considerable attention. Further, we also demonstrate how we can use it on timed automata and timed multi-stack pushdown systems (with boundedness restrictions).

7.2. Interrupt Timed Automata with Auxiliary Clocks and Parameters

Interrupt Timed Automata (ITA) are an expressive timed model, introduced to take into account interruptions according to levels. Due to this feature, this formalism is incomparable with Timed Automata. However several decidability results related to reachability and model checking have been obtained. In , we add auxiliary clocks to ITA, thereby extending its expressive power while preserving decidability of reachability. Moreover, we define a parametrized version of ITA, with polynomials of parameters appearing in guards and updates. While parametric reasoning is particularly relevant for timed models, it very often leads to undecidability results. We prove that various reachability problems, including robust reachability, are decidable for this model, and we give complexity upper bounds for a fixed or variable number of clocks, levels and parameters.

7.3. One-Counter Automata with Counter Observability

In a one-counter automaton (OCA), one can produce a letter from some finite alphabet, increment and decrement the counter by one, or compare it with constants up to some threshold. It is well-known that universality and language inclusion for OCAs are undecidable. In [14], we consider OCAs with counter observability: Whenever the automaton produces a letter, it outputs the current counter value along with it. Hence, its language is now a set of words over an infinite alphabet. We show that universality and inclusion for that model are PSPACE-complete, thus no harder than the corresponding problems for finite automata. In fact, by establishing a link with visibly one-counter automata, we show that OCAs with counter observability are effectively determinizable and closed under all boolean operations.

7.4. Diagnosis in Infinite-State Probabilistic Systems

In a recent work, we introduced four variants of diagnosability (FA, IA, FF, IF) in (finite) probabilistic systems (pLTS) depending whether one considers (1) finite or infinite runs and (2) faulty or all runs. We studied their relationship and established that the corresponding decision problems are PSPACE-complete. A key ingredient of the decision procedures was a characterisation of diagnosability by the fact that a random run almost surely lies in an open set whose specification only depends on the qualitative behaviour of the pLTS. In [12], we investigate similar issues for infinite pLTS. We first show that this characterisation still holds for FF-diagnosability but with a Gδ set instead of an open set and also for IF-and IA-diagnosability when pLTS are finitely branching. We also prove that surprisingly FA-diagnosability cannot be characterised in this way even in the finitely branching case. Then we apply our characterisations for a partially observable probabilistic extension of visibly pushdown automata (POpVPA), yielding EXPSPACE procedures for solving diagnosability problems. In addition, we establish some computational lower bounds and show that slight extensions of POpVPA lead to undecidability.

7.5. Accurate Approximate Diagnosability of Stochastic Systems

Diagnosis of partially observable stochastic systems prone to faults was introduced in the late nineties. Diagnosability, i.e. the existence of a diagnoser, may be specified in different ways: (1) exact diagnosability (called A-diagnosability) requires that almost surely a fault is detected and that no fault is erroneously claimed while (2) approximate diagnosability (called ε-diagnosability) allows a small probability of error when claiming a fault and (3) accurate approximate diagnosability (called AA-diagnosability) requires that this error threshold may be chosen arbitrarily small. In [11], we mainly focus on approximate diagnoses. We first refine the almost sure requirement about finite delay introducing a uniform version and showing that while it does not discriminate between the two versions of exact diagnosability this is no more the case in approximate diagnosis. Then we establish a complete picture for the decidability status of the diagnosability problems: (uniform) ε-diagnosability and uniform AA-diagnosability are undecidable while AA-diagnosability is decidable in PTIME, answering a longstanding open question.

7.6. Diagnosability of Repairable Faults

The diagnosis problem for discrete event systems consists in deciding whether some fault event occurred or not in the system, given partial observations on the run of that system. Diagnosability checks whether a correct diagnosis can be issued in bounded time after a fault, for all faulty runs of that system. This problem appeared two decades ago and numerous facets of it have been explored, mostly for permanent faults. It is known for example that diagnosability of a system can be checked in polynomial time, while the construction of a diagnoser is exponential. In [21], we examine the case of transient faults, that can appear and be repaired. Diagnosability in this setting means that the occurrence of a fault should always be detected in bounded time, but also before the fault is repaired. Checking this notion of diagnosability is proved to be PSPACE-complete. It is also shown that faults can be reliably counted provided the system is diagnosable for faults and for repairs.

7.7. Optimal constructions for active diagnosis

The task of diagnosis consists in detecting, without ambiguity, occurrence of faults in a partially observed system. Depending on the degree of observability, a discrete event system may be diagnosable or not. Active diagnosis aims at controlling the system in order to make it diagnosable. Solutions have already been proposed for the active diagnosis problem, but their complexity remains to be improved. In [8], we solve the active diagnosability decision problem and the active diagnoser synthesis problem, proving that (1) our procedures are optimal w.r.t. to computational complexity, and (2) the memory required for the active diagnoser produced by the synthesis is minimal. Furthermore, focusing on the minimal delay before detection, we establish that the memory required for any active diagnoser achieving this delay may be highly greater than the previous one. So we refine our construction to build with the same complexity and memory requirement an active diagnoser that realizes a delay bounded by twice the minimal delay.

7.8. Verification of parameterized communicating automata via split-width

In [16] study verification problems for distributed systems communicating via unbounded FIFO channels. The number of processes of the system as well as the communication topology are not fixed a priori. Systems are given by parameterized communicating automata (PCAs) which can be run on any communication topology of bounded degree, with arbitrarily many processes. Such systems are Turing powerful so we concentrate on under-approximate verification. We extend the notion of split-width to behaviors of PCAs. We show that emptiness, reachability and model-checking problems of PCAs are decidable when restricted to behaviors of bounded split-width. Reachability and emptiness are EXPTIME-complete, but only polynomial in the size of the PCA. We also describe several concrete classes of bounded split-width, for which we prove similar results.

7.9. Cyclic Ordering through Partial Orders

The orientation problem for ternary cyclic order relations has been attacked in the literature from combinatorial perspectives, through rotations, and by connection with Petri nets. In [7], we propose a two-fold characterization of orientable cyclic orders in terms of symmetries of partial orders as well as in terms of separating sets (cuts). The results are inspired by properties of non-sequential discrete processeses, but also apply to dense structures of any cardinality.

7.10. Predicting Traffic Load in Public Transportation Networks

This work is part of an ongoing effort to understand the dynamics of passenger loads in modern, multimodal transportation networks (TNs) and to mitigate the impact of perturbations, under the restrictions that the precise number of passengers in some point of the TN that intend to reach a certain destination (i.e. their distribution over different trip profiles) is unknown. In [29], we introduce an approach based on a stochastic hybrid automaton model for a TN that allows to compute how such probabilistic load vectors are propagated through the TN. In [23], [30], develop a computation strategy for forecasting the network's load a certain time in the future.

In [22], [28], we continue our work on perturbation analysis of multimodal transportation networks (TNs) by means of a stochastic hybrid automaton (SHA) model. We focus here on the approximate computation , in particular on the major bottleneck consisting in the high dimensionality of systems of stochastic differential balance equations (SDEs) that define the continuous passenger-flow dynamics in the different modes of the SHA model. In fact, for every pair of a mode and a station, one system of coupled SDEs relates the passenger loads of all discrete points such as platforms considered in this station, and all vehicles docked to it, to the passenger flows in between. In general, such an SDE system has many dimensions, which makes its numerical computation and thus the approximate computation of the SHA model intractable. We show how these systems can be canonically replaced by lower-dimensional ones, by decoupling the passenger flows inside every mode from one another. We prove that the resulting approximating passenger-flow dynamics converges to the original one, if the replacing set of balance equations set up for all decoupled passenger flows communicate their results among each other in vanishing time intervals.

For more information about the whole project, see [27].

7.11. Unfolding of Parametric Logical Regulatory Networks

In systems biology, models of cellular regulatory processes such as gene regulatory networks or signalling pathways are crucial to understanding the behaviour of living cells. Available biological data are however often insufficient for full model specification. In [18], we focus on partially specified models where the missing information is abstracted in the form of parameters. We introduce a novel approach to analysis of parametric logical regulatory networks addressing both sources of combinatoric explosion native to the model. First, we introduce a new compact representation of admissible parameters using Boolean lattices. Then, we define the unfolding of parametric regulatory networks. The resulting structure provides a partial- order reduction of concurrent transitions, and factorises the common transitions among the concrete models. A comparison is performed against state-of-the-art approaches to parametric model analysis.

7.12. Relationship between the Reprogramming Determinants of Boolean Networks and their Interaction Graph

In [24], we address the formal characterization of tar- gets triggering cellular trans-differentiation in the scope of Boolean net- works with asynchronous dynamics. Given two fixed points of a Boolean network, we are interested in all the combinations of mutations which allow to switch from one fixed point to the other, either possibly, or in- evitably. In the case of existential reachability, we prove that the set of nodes to (permanently) flip are only and necessarily in certain connected components of the interaction graph. In the case of inevitable reachabil- ity, we provide an algorithm to identify a subset of possible solutions.

7.13. D-SPACES: An Implementation of Declarative Semantics for Spatially Structured Information

We introduce in [17] D-SPACES, an implementation of constraint systems with space and extrusion operators. Constraint systems are algebraic models that allow for a semantic language-like representation of information in systems where the concept of space is a primary structural feature. We give this information mainly an epistemic interpretation and consider various agents as entities acting upon it. D-SPACES is coded as a c++11 library providing implementations for constraint systems, space functions and extrusion functions. The interfaces to access each implementation are minimal and thoroughly documented. D-SPACES also provides property-checking methods as well as an implementation of a specific type of constraint systems (a boolean algebra). This last implementation serves as an entry point for quick access and proof of concept when using these models. Furthermore, we offer an illustrative example in the form of a small social network where users post their beliefs and utter their opinions.

7.14. Belief, Knowledge, Lies and Other Utterances in an Algebra for Space and Extrusion

The notion of constraint system (cs) is central to declarative formalisms from concurrency theory such as process calculi for concurrent constraint programming (ccp). Constraint systems are often represented as lattices: their elements, called constraints, represent partial information and their order corresponds to entailment. Recently a notion of n-agent spatial cs was introduced to represent information in concurrent constraint programs for spatially distributed multi-agent systems. From a computational point of view a spatial constraint system can be used to specify partial information holding in a given agent's space (local information). From an epistemic point of view a spatial cs can be used to specify information that a given agent considers true (beliefs). Spatial constraint systems, however, do not provide a mechanism for specifying the mobility of information/processes from one space to another. Information mobility is a fundamental aspect of concurrent systems. In [6] we develop the theory of spatial constraint systems with operators to specify information and processes moving from a space to another. We shall investigate the properties of this new family of constraint systems and illustrate their applications. From a computational point of view the new operators provide for process/information extrusion, a central concept in formalisms for mobile communication. From an epistemic point of view extrusion corresponds I to a notion we shall call utterance; a piece of information that an agent communicate to others but that may be inconsistent with the agent's beliefs. Utterances can then be used to express instances of epistemic notions such as hoaxes or intentional lies which are common place in social media. Spatial constraint system can express the epistemic notion of belief by means of space functions that specify local information. We shall also show that spatial constraint can also express the epistemic notion of knowledge by means of a derived spatial operator that specifies global information.

7.15. Goal-Driven Unfolding of Petri Nets

Unfoldings provide an efficient way to avoid the state-space explosion due to interleavings of concurrent transitions when exploring the runs of a Petri net. The theory of adequate orders allows one to define finite prefixes of unfoldings which contain all the reachable markings. In this paper we are interested in reachability of a single given marking, called the goal. In [26], We propose an algorithm for computing a finite prefix of the unfolding of a 1-safe Petri net that preserves all minimal configurations reaching this goal. Our algorithm combines the unfolding technique with on-the-fly model reduction by static analysis aiming at avoiding the exploration of branches which are not needed for reaching the goal. We present some experimental results.

PARSIFAL Project-Team

7. New Results

7.1. Linear rewriting systems for Boolean logic

Participant: Lutz Straßburger.

Last year's result on the nonexistence of a complete linear term rewriting system for propositional logic [53] has been generalized and some applications to proof theory have been invesigated. For example, we have found that the medial rule which plays a central role in deep inference systems is canonical in a strong sense: It is minimal, and every rule that reduce contraction to an atomic form is indeed derivable via medial. This is published in [15] (joint work with Anupam Das).

7.2. Non-crossing Tree Realizations of Ordered Degree Sequences

Participant: Lutz Straßburger.

We investigate the enumeration of non-crossing tree realizations of integer sequences, and we consider a special case in four parameters, that can be seen as a four-dimensional tetrahedron that generalizes Pascal's triangle and the Catalan numbers. This work is motivated by the study of ambiguities arising in the parsing of natural language sentences using categorial grammars. This is joint work with Laurent Méhats and published in [31].

7.3. Focusing for Nested Sequents

Participants: Kaustuv Chaudhuri, Sonia Marin, Lutz Straßburger.

Focusing is a general technique for transforming a sequent proof system into one with a syntactic separation of non-deterministic choices without sacrificing completeness. This not only improves proof search, but also has the representational benefit of distilling sequent proofs into synthetic normal forms. We have shown how to apply the focusing technique to nested sequent calculi, a generalization of ordinary sequent calculi to tree-like instead of list-like structures. We thus improve the reach of focusing to the most commonly studied modal logics, the logics of the modal S5 cube. Among our key contributions is a focused cut-elimination theorem for focused nested sequents. This is published in [25].

Then we further extend our results to intuitionistic nested sequents, which can capture all the logics of the intuitionistic S5 cube in a modular fashion. We obtained an internal cut-elimination procedure for the focused system which in turn is used to show its completeness. This is published in [26]

7.4. Combining inference systems: a generalization of Nelson-Oppen and MCSAT

Participant: Stéphane Graham-Lengrand.

Nelson-Oppen [79] and Model-Constructing Satisfiability (MCSAT) [89], [65] are two methodologies that allow the reasoning mechanisms of different theories to collaborate, in order to tackle hybrid problems. While these methodologies are often used and implemented for the practical applications of Automated Reasoning, their rather sophisticated foundations are traditionally explained in terms of model theory. SRI International pioneered some work providing such methodologies with new and more general foundations in terms of *inference systems* [57], closer to proof theory and to Parsifal's research. The more recent MCSAT methodology was not captured, more generally lacked any kind of theorem about the generic combination of arbitrary theories, and was also thought to be incompatible with the Nelson-Oppen approach, so that SMT-solvers are either working with one methodology or the other, unable to get the best of both worlds.

In 2016 we designed a combination methodology, based on *inference systems*, that supersedes both Nelson-Oppen and MCSAT [34]. We showed its soundness and completeness, and identified for this the properties that the theories to combine are required to satisfy. This generalized MCSAT with the generic combination mechanism that it lacked, and showed that it is perfectly compatible with the Nelson-Oppen methodology, which can now cohabit within the same solver.

7.5. Linear lambda terms as invariants of rooted trivalent maps

Participant: Noam Zeilberger.

Recent studies of the combinatorics of linear lambda calculus have uncovered some unexpected connections to the old and well-developed theory of graphs embedded on surfaces (also known as "maps") [47], [87], [88]. In [19], we aimed to give a simple and conceptual account for one of these connections, namely the correspondence (originally described by Bodini, Gardy, and Jacquot [47]) between α-equivalence classes of closed linear lambda terms and isomorphism classes of rooted trivalent maps on compact oriented surfaces without boundary. One immediate application of this new account was a characterization of trivalent maps which are *bridgeless* (in the graph-theoretic sense of having no disconnecting edge) as linear lambda terms with no closed proper subterms. In turn, this lead to a surprising but natural reformulation of the Four Color Theorem as a statement about typing in lambda calculus.

7.6. A bifibrational reconstruction of Lawvere's presheaf hyperdoctrine

Participant: Noam Zeilberger.

In joint work with Paul-André Melliès, we have been investigating the categorical semantics of type refinement systems, which are type systems built "on top of" a typed programming language to specify and verify more precise properties of programs. The fibrational view of type refinement we have been developing (cf. [72]) is closely related to the categorical perspective on first-order logic introduced by Lawvere [66], but with some important conceptual and technical differences that provide an opportunity for reflection. For example, Lawvere's axiomatization of first-order logic (his theory of so-called "hyperdoctrines") was based on the idea that existential and universal quantification can be described respectively as left and right adjoints to the operation of substitution, this giving rise to a family of *adjoint triples* $\Sigma_f \dashv \mathcal{P}_f \dashv \Pi_f$ (one such triple for every function $f : A \to B$). On the other hand, a bifibration only induces a family of *adjoint pairs* $\mathsf{push}_f \dashv \mathsf{pull}_f$ (again, one such pair for every $f : A \to B$). In [33], we resolved this and other apparent mismatches by applying ideas inspired by the semantics of linear logic and the shift from the cartesian closed category $\mathbf{Set}$ to the symmetric monoidal closed category $\mathbf{Rel}$. Two other applications of our analysis include an axiomatic treatment of *directed* equality predicates (which can be modelled as "hom" presheaves, realizing an early vision of Lawvere), as well as a simple calculus of string diagrams that is highly reminiscent of C. S. Peirce's "existential graphs" for predicate logic.

7.7. Towards a link between CPS and focusing

Participant: Matthias Puech.

Continuation-passing style translations make a functional program more explicit by sequentializing its computations and reifying its control. They have been used as an intermediate language in many compilers. They are also understood as classical-to-intuitionistic proof embedding (so-called double negation translations). Matthias Puech studied a novel correspondence between CPS and focusing: to each CPS transform corresponds a focused proof system that is identifiable as a particular polarization of classical statements. Since, after Miller's and others work, we know the full design space of focused sequent calculi, we expect to understand the full design space of CPS translation.

The first step of this goal is to study the syntax and typing of variants of the CPS translation. Puech designed and implemented in OCaml a compacting, optimizing CPS translation, while using OCaml's type system to verify that it maps well-typed terms to well-typed terms in a tightly restricted syntactical form (the "typeful" approach to formalization) [82]. The resulting type system is in Curry-Howard isomorphism with a weakly focused proof system: LJQ.

7.8. Proof Checking and Logic Programming

Participants: Roberto Blanco, Tomer Libal, Dale Miller, Marco Volpe.

In a world where trusting software systems is increasingly important, formal methods and formal proofs can help provide some basis for trust. Proof checking can help to reduce the size of the *trusted base* since we do not need to trust an entire theorem prover: instead, we only need to trust a (smaller and simpler) proof checker. Many approaches to building proof checkers require embedding within them a full programming language. In most modern proof checkers and theorem provers, that programming language is a functional programming language, often a variant of ML. In fact, aspects of ML (e.g., strong typing, abstract data types, and higher-order programming) were designed to make ML a trustworthy "meta-language" for checking proofs. While there is considerable overlap between logic programming and proof checking (e.g., both benefit from unification, backtracking search, efficient term structures, etc), the discipline of logic programming has, in fact, played a minor role in the history of proof checking. Miller has been pushing the argument that logic programming can have a major role in the future of this important topic [18]. Many aspects of the ProofCert project are based on this perspective that logic programming techniques and methods can have significant utility within proof checking. This perspective stands in constrast to the work on the Dedukti proof checking framework [44] where functional programming principles are employed for proof checking.

7.9. Proof Certificates for First-Order Equational Logic

Participants: Dale Miller, Zakaria Chihani.

The kinds of inference rules and decision procedures that one writes for proofs involving equality and rewriting are rather different from proofs that one might write in first-order logic using, say, sequent calculus or natural deduction. For example, equational logic proofs are often chains of replacements or applications of oriented rewriting and normal forms. In contrast, proofs involving logical connectives are trees of introduction and elimination rules. Chihani and Miller have shown [13] how it is possible to check various equality-based proof systems with a programmable proof checker (the *kernel* checker) for first-order logic. That proof checker's design is based on the implementation of *focused proof search* and on making calls to (user-supplied) *clerks and experts* predicates that are tied to the two phases found in focused proofs. This particular design is based on the work of Chihani, Miller, and Renaud [14].

The specification of these clerks and experts provide a formal definition of the structure of proof evidence and they work just as well in the equational setting as in the logic setting where this scheme for proof checking was originally developed. Additionally, executing such a formal definition on top of a kernel provides an actual proof checker that can also do a degree of proof reconstruction. A number of rewriting based proofs have been defined and checked in this manner.

7.10. Extended Pattern Unification

Participants: Tomer Libal, Dale Miller.

Unification is a central operation in the construction of a range of computational logic systems based on first-order and higher-order logics. First-order unification has a number of properties that dominates the way it is incorporated within such systems. In particular, first-order unification is decidable, unary, and can be performed on untyped term structures. None of these three properties hold for full higher-order unification: unification is undecidable, unifiers can be incomparable, and term-level typing can dominate the search for unifiers. The so-called *pattern* subset of higher-order unification was designed to be a small extension to first-order unification that respected the basic laws governing λ-binding (the equalities of α, β, and η-conversion) but which also satisfied those three properties. While the pattern fragment of higher-order unification has been popular in various implemented systems and in various theoretical considerations, it is too weak for a number of applications. Libal and Miller [28] have defined an extension of pattern unification that is motivated by some existing applications and which satisfies these three properties. The main idea behind their extension is that the arguments to a higher-order, free variable can be more than just distinct bound variables: they can also be terms constructed from (sufficient numbers of) such variables using term constructors and where no

argument is a subterm of any other argument. This extension to pattern unification satisfies the three properties mentioned above. R. Blanco is currently adding this extended unification to the Abella theorem prover.

7.11. Focused proofs for modal logics

Participants: Tomer Libal, Sonia Marin, Dale Miller, Marco Volpe.

Several deductive formalisms (e.g., sequent, nested sequent, labeled sequent, hypersequent calculi) have been used in the literature for the treatment of modal logics, and some connections between these formalisms are already known. Marin, Miller, and Volpe [30] have propose a general framework, which is based on a focused version of the labeled sequent calculus by Negri [78], augmented with some parametric devices allowing to restrict the set of proofs. By properly defining such restrictions and by choosing an appropriate polarization of formulas, one can obtain different, concrete proof systems for the modal logic K and for its extensions by means of geometric axioms. The expressiveness of the labeled approach and the control mechanisms of focusing allow a clean emulation of a range of existing formalisms and proof systems for modal logic. These results make it possible to write Foundational Proof Certificate definitions of common modal logic proof systems.

7.12. Preserving differential privacy under finite-precision semantics

Participant: Dale Miller.

(Joint work with Ivan Gazeau and Catuscia Palamidessi). The approximation introduced by finite-precision representation of continuous data can induce arbitrarily large information leaks even when the computation using exact semantics is secure. Such leakage can thus undermine design efforts aimed at protecting sensitive information. Gazeau, Miller, and Palamidessi [16] have applied differential privacy—an approach to privacy that emerged from the area of statistical databases—to this problem. In their approach, privacy is protected by the addition of noise to a true (private) value. To date, this approach to privacy has been proved correct only in the ideal case in which computations are made using an idealized, infinite-precision semantics. An analysis of implementation levels, where the semantics is necessarily finite-precision, i.e. the representation of real numbers and the operations on them are rounded according to some level of precision. In general there are violations of the differential privacy property but a limited (but, arguably, totally acceptable) variant of the property can be used insteand, under only a minor degradation of the privacy level. In fact, two cases of noise-generating distributions can be employed: the standard Laplacian mechanism commonly used in differential privacy, and a bivariate version of the Laplacian recently introduced in the setting of privacy-aware geolocation.

7.13. Certification of Prefixed Tableau Proofs for Modal Logic

Participants: Tomer Libal, Marco Volpe.

This work [29] describes the theory and implementation of a proof checker for tableau theorem provers for modal logics. The tool supports proofs in both the traditional tableau format as well as the free variable variant. The implentation can be found at https://github.com/proofcert/checkers under the gandalf2016 branch.

7.14. Towards a Substitution Tree Based Index for Higher-order Resolution Theorem Provers

Participant: Tomer Libal.

First-order resolution theorem provers depend on efficient data structures for redundancy elimination. These data structures do not exist for higher-order resolution theorem provers. In [32] we discuss a new approach to this problem. (Joint work with Alexander Steen).

7.15. Open Call-by-Value

Participant: Beniamino Accattoli.

Functional programming languages are often based on the call-by-value λ-calculus, whose elegant theory relies on weak evaluation and closed terms, that are natural hypotheses in the study of programming languages. To model proof assistants, however, strong evaluation and open terms are required, and it is well known that the operational semantics of call-by-value becomes problematic in this case. In this joint work with Giulio Guerrieri we studied the intermediate setting—that we call Open Call-by-Value—of weak evaluation with open terms, on top of which Gregoire and Leroy designed the abstract machine of Coq. Various calculi for Open Call-by-Value already exist, each one with its pros and cons. We did a detailed comparative study of the operational semantics of four of them, coming from different areas such as the study of abstract machines, denotational semantics, linear logic proof nets, and sequent calculus. We showed that these calculi are all equivalent from a termination point of view, justifying the slogan Open Call-by-Value. The work has been published in the proceedings of the international conference APLAS 2016 [22].

7.16. A Reasonable Abstract Machine for the Strong λ-Calculus

Participant: Beniamino Accattoli.

We provided a new proof that the strong λ-calculus is a reasonable computational model. The original proof is by B. Accattoli and H. Dal Lago uses a calculus with explicit substitutions while the new one relies on a new sophisticated abstract machine, the Useful MAM. The work has been published in the proceeding of the international conference WoLLIC 2016 [21].

7.17. Space-efficient Acyclicity Constraints

Participant: Taus Brock-Nannestad.

Acyclicity constraints can be used to encode a large variety of useful constraints on graphs. The basic constraint itself can be encoded in terms of simpler constraints (e.g. integer linear constraints) in a straightforward and intuitive way, associating to each vertex of the (fixed) input graph a variable with domain linear in the size of the graph. For large graphs, this quickly becomes inefficient.

In [24], we show that in the case of planar graphs, a more efficient encoding (using a two-valued variable per vertex) is possible.

7.18. Exp-log normal form of types and the axioms for η-equality of the λ-calculus with sums

Participant: Danko Ilik.

In the presence of sum types, the λ-calculus has but one implemented (and incomplete) heuristic for deciding $\beta\eta$-equality of terms, in spite of a dozen of meta-theoretic works showing that the equality is decidable.

In the work discussed here, we first used the exp-log decomposition of the arrow type—inspired from the analytic transformation $a^b = \exp(b \times \log a)$—to obtain a type normal form for the type languages $\{\rightarrow, \times, +\}$. We then made a quotient of the $\beta\eta$-equality of terms modulo the terms coerced into their representation at the exp-log normal form of their type. This allows to obtain a *simplification* of the so far standard axioms for $\beta\eta$-equality.

Moreover, we provided a Coq implementation of a heuristic decision procedure for this equality. Although a heuristic, this implementation manages to tackle examples of equal terms that need a complex program analysis in the only previously implemented heuristic of Vincent Balat.

This work is described in a paper accepted for presentation at POPL 2017, [27].

7.19. Invertible-rule-free sequent calculi and an intuitionistic arithmetical hierarchy

Participants: Taus Brock-Nannestad, Danko Ilik.

In sequent calculi, proof rules can be divided into two groups: invertible (asynchronous) proof rules and non-invertible (synchronous) proof rules. Even in focusing sequent calculi the two groups of rules are present, albeit grouped together in synthetic rules (we speak of the synchronous and asynchronous phase).

In this work, we used the exp-log decomposition (described above) in the context of logic in order to obtain a version of sequent caclulus which contains synchronous rules only, a first such formalism for intuitionistic logic.

We extended the picture from the setting of propositional to the one of first-order intuitionistic logic, where the exp-log decomposition provided us with an intuitionistic hierarchy of formulas analogous to the classical arithmetical hierachy; although the classical arithmetical hierarchy exists since the 1920s, a correspondingly versatile notion for intuitionistic logic has been elusive up to this day.

This work is described in the manuscript [37], submitted to an academic journal.

SPECFUN Project-Team

6. New Results

6.1. Formally certified computation of definite integrals

Assia Mahboubi and Thomas Sibut-Pinote, in collaboration with Guillaume Melquiond (Toccata), have developed a Coq library for the computation of intervals approximating the value of definite integrals for elementary mathematical functions. This library provides an automated tool which builds automatically a formal proof of the correctness of the output, that is: a formal proof that the interval contains the mathematical values and a formal proof of the integrability of the input function on the input interval. A description of this work was published in the proceeding of the ITP 2016 conference [13]. An extension to domains including singularities of the integrand is in progress.

6.2. Real closed fields

Assia Mahboubi has worked with Henri Lombardi (Université de Franche Comté) on a constructive axiomatization of real closed fields. For this purpose, they have proposed an equational theory based on virtual roots and close to the classical notion of local real closed rings. This is a first step toward a constructive understanding of o-minimal structures. This work has been accepted for publication in Contemporary Mathematics [19].

6.3. Combinatorial walks with small steps in the quarter plane

Alin Bostan and Frédéric Chyzak, together with Mark van Hoeij (Florida State University), Manuel Kauers (Johannes Kepler University), and Lucien Pech (former intern), have applied their algorithms on special functions to generate complete, quantitative results in the enumerative theory of combinatorial walks with small steps in the quarter plane [2]. They gave the first proof that differential equations conjectured years ago by Bostan and Kauers are indeed satisfied by the corresponding generating functions. They also obtained explicit hypergeometric expressions for the latter, and could provably determine which of the generating functions are transcendental or algebraic.

6.4. Multiple binomial sums

Multiple binomial sums form a large class of multi-indexed sequences, closed under partial summation, which contains most of the sequences obtained by multiple summation of products of binomial coefficients and also all the sequences with algebraic generating function. Alin Bostan and Pierre Lairez, together with Bruno Salvy (Inria and ENS Lyon), have studied in [5] the representation of the generating functions of binomial sums by integrals of rational functions. The outcome is twofold. Firstly, we show that a univariate sequence is a multiple binomial sum if and only if its generating function is the diagonal of a rational function. Secondly, we propose algorithms that decide the equality of multiple binomial sums and that compute recurrence relations for them. In conjunction with geometric simplifications of the integral representations, this approach behaves well in practice. The process avoids the computation of certificates and the problem of the appearance of spurious singularities that afflicts discrete creative telescoping, both in theory and in practice.

6.5. Algebraic diagonals and walks

The diagonal of a multivariate power series F is the univariate power series $\mathsf{Diag}F$ generated by the diagonal terms of F. Diagonals form an important class of power series; they occur frequently in number theory, theoretical physics and enumerative combinatorics. In [35], Alin Bostan and Louis Dumont, together with Bruno Salvy (Inria and ENS Lyon), have studied algorithmic questions related to diagonals in the case where F is the Taylor expansion of a bivariate rational function. It is classical that in this case $\mathsf{Diag}F$ is an algebraic function. We propose an algorithm that computes an annihilating polynomial for $\mathsf{Diag}F$. We give a precise bound on the size of this polynomial and show that generically, this polynomial is the minimal polynomial and that its size reaches the bound. The algorithm runs in time quasi-linear in this bound, which grows exponentially with the degree of the input rational function. We then address the related problem of enumerating directed lattice walks. The insight given by our study leads to a new method for expanding the generating power series of bridges, excursions and meanders. We show that their first N terms can be computed in quasi-linear complexity in N, without first computing a very large polynomial equation. An extended version of this work is presented in [3].

6.6. A human proof of the Gessel conjecture

Counting lattice paths obeying various geometric constraints is a classical topic in combinatorics and probability theory. Many recent works deal with the enumeration of 2-dimensional walks with prescribed steps confined to the positive quadrant. A notoriously difficult case concerns the so-called *Gessel walks*: they are planar walks confined to the positive quarter plane, that move by unit steps in any of the following directions: West, North-East, East and South-West. In 2001, Ira Gessel conjectured a closed-form expression for the number of such walks of a given length starting and ending at the origin. In 2008, Kauers, Koutschan and Zeilberger gave a computer-aided proof of this conjecture. The same year, Bostan and Kauers showed, using again computer algebra tools, that the trivariate generating function of Gessel walks is algebraic. Alin Bostan, together with Irina Kurkova (Univ. Paris 6) and Kilian Raschel (CNRS and Univ. Tours), have proposed in [4] the first "human proofs" of these results. They are derived from a new expression for the generating function of Gessel walks in terms of special functions.

6.7. Enumeration of 3-dimensional lattice walks confined to the positive octant

Small step walks in 2D are by now quite well understood, but almost everything remains to be done in higher dimensions. Alin Bostan, together with Mireille Bousquet-Mélou (CNRS and Univ. Bordeaux), Manuel Kauers (Johannes Kepler Univ.) and Stephen Melczer (Univ. of Waterloo and ENS Lyon), have explored in [1] the classification problem for 3-dimensional walks with unit steps confined to the positive octant. The first difficulty is their number: there are 11 074 225 cases (instead of 79 in dimension 2). In our work, we focused on the 35 548 that have at most six steps. We applied to them a combined approach, first experimental and then rigorous. Among the 35 548 cases, we first found 170 cases with a finite group; in the remaining cases, our experiments suggest that the group is infinite. We then rigorously proved D-finiteness of the generating series in all the 170 cases, with the exception of 19 intriguing step sets for which the nature of the generating function still remains unclear. In two challenging cases, no human proof is currently known, and we derived computer-algebra proofs, thus constituting the first proofs for those two step sets.

6.8. Computation of the similarity class of the p-curvature

The p-curvature of a system of linear differential equations in positive characteristic p is a matrix that measures how far the system is from having a basis of polynomial solutions. Alin Bostan, together with Xavier Caruso (CNRS and Univ. Rennes) and Éric Schost (Univ. Waterloo), have showed in [10] that the similarity class of the p-curvature can be determined without computing the p-curvature itself. More precisely, we have designed an algorithm that computes the invariant factors of the p-curvature in time quasi-linear in $\sqrt{p}$. This is much less than the size of the p-curvature, which is generally linear in p. The new algorithm allowed to answer a question originating from the study of the Ising model in statistical physics.

6.9. Efficient algorithms for mixed creative telescoping

Creative telescoping is a powerful computer algebra paradigm –initiated by Doron Zeilberger in the 90's– for dealing with definite integrals and sums with parameters. Alin Bostan and Louis Dumont, together with Bruno Salvy (Inria and ENS Lyon), have addressed in [12] the mixed continuous–discrete case, and have focussed on the integration of bivariate hypergeometric-hyperexponential terms. We have designed a new creative telescoping algorithm operating on this class of inputs, based on a Hermite-like reduction procedure. The new algorithm has two nice features: it is efficient and it delivers, for a suitable representation of the input, a minimal-order telescoper. Its analysis reveals tight bounds on the sizes of the telescoper it produces.

6.10. Fast computation of the Nth term of an algebraic series over a finite prime field

Alin Bostan and Philippe Dumas, together with Gilles Christol (IMJ), have addressed in [11] the question of computing one selected term of an algebraic power series. In characteristic zero, the best algorithm currently known for computing the Nth coefficient of an algebraic series uses differential equations and has arithmetic complexity quasi-linear in $\sqrt{N}$. We show that over a prime field of positive characteristic p, the complexity can be lowered to $O(\log N)$. The mathematical basis for this dramatic improvement is a classical theorem stating that a formal power series with coefficients in a finite field is algebraic if and only if the sequence of its coefficients can be generated by an automaton. We revisit and enhance two constructive proofs of this result for finite prime fields. The first proof uses Mahler equations, whose sizes appear to be prohibitively large. The second proof relies on diagonals of rational functions; we turn it into an efficient algorithm, of complexity linear in $\log N$ and quasi-linear in p.

6.11. Formal methods for cryptocurrencies

Georges Gonthier and Thomas Sibut-Pinote, along with a team of researchers from Microsoft Research and Inria, participated in a hackathon internal to Microsoft Research with the goal to apply formal methods to the verification of the smart contracts involved in the Ethereum platform. They outlined a framework to analyze and verify both the runtime safety and the functional correctness of Ethereum contracts by translation to F*, a functional programming language aimed at program verification. This work was published in the proceedings of the PLAS 2016 conference [9].

6.12. Computing solutions of linear Mahler equations

Mahler equations relate evaluations of the same function f at iterated bth powers of the variable. They arise in particular in the study of automatic sequences and in the complexity analysis of divide-and-conquer algorithms. Recently, the problem of solving Mahler equations in closed form has occurred in connection with number-theoretic questions. A difficulty in the manipulation of Mahler equations is the exponential blow-up of degrees when applying a Mahler operator to a polynomial. In [17], Frédéric Chyzak and Philippe Dumas, together with Thomas Dreyfus (Université Claude Bernard Lyon 1) and Marc Mezzarobba (visiting scientist from UPMC), have presented algorithms for solving linear Mahler equations for series, polynomials, and rational functions, and have obtained polynomial-time complexity under a mild assumption.

6.13. Formal solutions of singularly perturbed linear differential systems

Suzy Maddah, together with Boulay Barkatou (Université de Limoges), has obtained algorithms for computing formal invariants of singularly-perturbed linear differential systems [20].

TOCCATA Project-Team

7. New Results

7.1. Deductive Verification

A bit-vector library for deductive verification. C. Fumex and C. Marché developed a new library for bit-vectors, in Why3 and SPARK. This library is rich enough for the formal specification of functional behavior of programs that operate at the level of bits. It is also designed to exploit efficiently the support for bit-vectors built-in in some SMT solvers. This work is done in the context of the ProofInUse joint laboratory. The SPARK front-end of Why3, for the verification of Ada programs, is extended to exploit this new bit-vector theory. Several cases studies are conducted: efficient search for rightmost bit of a bit-vector, efficient computation of the number of bits set to 1, efficient solving of the n-queens problem. At the level of SPARK, a program inspired from some industrial code (originally developed in C par J. Gerlach, Fraunhofer FOKUS Institute, Germany and partially proved with Frama-C and Coq) is specified in SPARK and proved with automatic solvers only. A paper on that library together with the way it is connected with the built-in support for bitvectors in SMT solver was presented at the NASA Formal methods Conference [24]. The support for bit-vectors is distributed with SPARK since 2015, and SPARK users already reported that several verification conditions, that couldn't be proved earlier, are now proved automatically.

Counterexamples from proof failures. D. Hauzar and C. Marché worked on counterexample generation from failed proof attempts. They designed a new approach for generating potential counterexamples in the deductive verification setting, and implemented in Why3. When the logic goal generated for a given verification condition is not shown unsatisfiable by an SMT solvers, some solver can propose a model. By carefully reverting the transformation chain (from an input program through the VC generator and the various translation steps to solvers), this model is turned into a potential counterexample that the user can exploit to analyze why its original code is not proved. The approach is implemented in the chain from Ada programs through SPARK, Why3, and SMT solvers CVC4 and Z3. This work is described in a research report [35] and a paper was rpesented at the SEFM Conference [25]. The work on the implementation was continued by S. Dailler. It was considered robust enough to be distributed in the release Pro 16 of SPARK.

Static versus dynamic verification. C. Marché, together with Y. Moy from AdaCore, J. Signoles and N. Kosmatov from CEA-LIST, wrote a survey paper about the design of the specification languages of Why3 and its front-ends Frama-C and SPARK. The choices made when designing these specification languages differ significantly, in particular with respect to the executability of specifications. The paper reviews these differences and the issues that result from these choices. The paper also emphasizes two aspects where static and dynamic aspects of the specifications play an important role: the specific feature of *ghost code*, and the techniques that help users understand why static verification fails. This paper was presented at the Isola Symposium [26].

Higher-Order Representation Predicates. A. Charguéraud investigated how to formalize in Separation Logic representation predicates for describing mutable container data structures that store mutable elements that are themselves described using representation predicates. (In Separation Logic, representation predicates are used to describe mutable data structures, by establishing a relationship between the entry point of the structure, the piece of heap over which this structure spans, and the logical model associated with the structure.) The solution proposed, based on "higher-order representation predicates", allows for concise specifications of such containers. A. Charguéraud has published a paper presenting, through a collection of practical examples, solutions to the challenges associated with verification proofs based on higher-order representation predicates [19].

Temporary Read-Only Permissions for Separation Logic A. Charguéraud and François Pottier (Inria Paris) have developed an extension of Separation Logic with temporary read-only permissions. This mechanism allows to temporarily convert any assertion (or "permission") to a read-only form. Unlike with fractional permissions, no accounting is required: the proposed read-only permissions can be freely duplicated and discarded. Where mutable data structures are temporarily accessed only for reading, the proposed read-only permissions enable more concise specifications and proofs. All the metatheory is verified in Coq. An article has been submitted to a conference [20].

Reasoning About Iteration. J.-C. Filliâtre and M. Pereira proposed a new approach to the problem of specifying iteration, verifying iterators (such as cursors or higher-order functions), and using iterators. The idea is to characterize the sequence of elements enumerated so far, and only those. The proof methodology is modular, iterator implementations and clients being verified independently of each other. The proposed method is validated experimentally in Why3. This work has been published first at JFLA 2016 [33] and then at NFM 2016 [22]. A journal version of this work is under submission.

Defunctionalization for proving higher-order programs. J.-C. Filliâtre and M. Pereira proposed a new approach to the verification of higher-order programs, using the technique of defunctionalization, that is, the translation of first-class functions into first-order values. This is an early experimental work, conducted on examples only within the Why3 system. This work has been published at JFLA 2017 [30].

A Type System for Deductive Verification. J.-C. Filliâtre, L. Gondelman, and A. Paskevich proposed a practical method to track pointer aliases statically in a large family of computer programs. Their approach relies on a special type system with singleton regions and effects which both can be inferred automatically, without requiring additional user annotations. This kind of static analysis is important for deductive program verification, since it allows us to construct verification conditions using the traditional rules in the spirit of Hoare and Dijkstra, without recurring to more sophisticated solutions (memory models, separation logic) which incur additional complexity both for a user and a verification tool. The proposed method is implemented in Why3 and described in a technical report [37].

Ghost Code. J.-C. Filliâtre, L. Gondelman, and A. Paskevich published a paper on a general approach to the concept of ghost code in the journal of *Formal Methods in System Design* [14]. Ghost code is a subset of program code that serves the purposes of specification and verification: it can be erased from the program without affecting its result. This work forms the basis of the support for ghost code in Why3. This work is an extended version of the paper presented at the 26th International Conference on Computer Aided Verification (CAV) in 2014.

7.2. Automated Reasoning

Decision Procedures via Axiomatizations with Triggers. C. Dross, A. Paskevich, J. Kanig and S. Conchon published a paper in the *Journal of Automated Reasoning* [13] about integration of first-order axiomatizations with triggers as decision procedures in an SMT solver. This work extends a part of C. Dross PhD thesis [83]. A formal semantics of the notion of trigger is presented, with a general setting to show how a first-order axiomatization with triggers can be proved correct, complete, and terminating. An extended DPLL(T) algorithm can then integrate such an axiomatization with triggers, as a decision procedure for the theory it defines.

Lightweight Approach for Declarative Proofs. M. Clochard designed an extension of first-order logic, for describing reasoning steps needed to discharge a proof obligation. The extension is under the form of two new connectives, called proof indications, that allow the user to encode reasoning steps inside a logic formula. This extension makes possible to use the syntax of formulas as a proof language. The approach was presented at the JFLA conference [29] and implemented in Why3. It brings a lightweight mechanism for declarative proofs in an environment like Why3 where provers are used as black boxes. Moreover, this mechanism restricts the scope of auxiliary lemmas, reducing the size of proof obligations sent to external provers.

7.3. Certification of Algorithms, Languages, Tools and Systems

Case study: Matrix Multiplication. M. Clochard, L. Gondelman and M. Pereira wrote a paper describing a complete solution for the first challenge of the VerifyThis 2016 competition held at the 18th ETAPS Forum, where they obtain the award for the best student team. Two variants for the multiplication of matrices are presented and proved: a naive version using three nested loops and Strassen's algorithm. To formally specify the two multiplication algorithms, they developed a new Why3 theory of matrices, and they applied a reflection methodology to condut some of the proofs. This work was presented at the VSTTE Conference [21]. An extended version that considers arbitrary rectangular martices instead of square ones is in preparation. The development is available at http://toccata.lri.fr/gallery/verifythis_2016_matrix_multiplication.en.html.

Case study: Koda-Ruskey's algorithm for generating ideals of a forest. J.-C. Filliâtre and M. Pereira presented the first formal proof of an implementation of Koda and Ruskey's algorithm (an algorithm for generating all ideals of a forest poset as a Gray code) at VSTTE 2016 [23]. The proof is conducted within the Why3 system and is mostly automatic.

The Lax-Milgram Theorem. S. Boldo, F. Clément, F. Faissole, V. Martin, and M. Mayero have worked on a Coq formal proof of the Lax–Milgram theorem. The Finite Element Method is a widely-used method to solve numerical problems coming for instance from physics or biology. To obtain the highest confidence on the correction of numerical simulation programs implementing the Finite Element Method, one has to formalize the mathematical notions and results that allow to establish the sound-ness of the method. The Lax–Milgram theorem may be seen as one of those theoretical cornerstones: under some completeness and coercivity assumptions, it states existence and uniqueness of the solution to the weak formulation of some boundary value problems. This article presents the full formal proof of the Lax–Milgram theorem in Coq. It requires many results from linear algebra, geometry, functional analysis , and Hilbert spaces. This has been published at the 6th ACM SIGPLAN Conference on Certified Programs and Proofs (CPP 2017) [18].

ALEA library extended with continuous datatypes The ALEA library uses a monadic construction to formalize discrete measure theory. F. Faissole and B. Spitters proposed to extend it to continuous datatypes. They used both synthetic topology and homotopy type theory to achieve the formalization. This work is presented at the Workshop on Coq for Programming Languages [32].

Case study: Strongly Connected Components of a Graph R. Chen and J.-J. Lévy designed a formal proof of Tarjan's algorithm for computing the strongly connected component of a directed graph. The proof is conducted using Why3. This work is presented at the JFLA conference [28]. This case study is part of a larger set of case studies on algorithms on graphs http://pauillac.inria.fr/~levy/why3/.

Case study: Unix Pathname Resolution R. Chen, M. Clochard and C.-Marché designed a formal proof of an algorithm for resolving a pathname in Unix file systems. The proof is conducted using Why3 [34]. This case study is part of the CoLiS project.

7.4. Floating-Point and Numerical Programs

-

Interval arithmetic and Taylor models. É. Martin-Dorel and G. Melquiond have worked on integrating the CoqInterval and CoqApprox libraries into a single package. The CoqApprox library is dedicated to computing verified Taylor models of univariate functions so as to compute approximation errors. The CoqInterval library reuses this work to automatically prove bounds on real-valued expressions. A large formalization effort took place during this work, so as to get rid of all the holes remaining in the formal proofs of CoqInterval. It was also the chance to perform a comparison between numerous decision procedures dedicated to proving nonlinear inequalities involving elementary functions. This work has been published in the *Journal of Automated Reasoning* [15].

Interval arithmetic and univariate integrals. A. Mahboubi, G. Melquiond, and T. Sibut-Pinote have extended the CoqInterval library with support for definite univariate integrals. The library is now able to automatically and formally verify bounds on the value of integrals by computing rigorous polynomial approximations of integrands. This work has been presented at the 7th International Conference on Interactive Theorem Proving [27].

Robustness of 2Sum and Fast2Sum. S. Boldo, S.Graillat, and J.-M. Muller have worked on the 2Sum and Fast2Sum algorithms, that are important building blocks in numerical computing. They are used (implicitely or explicitely) in many compensated algorithms or for manipulating floating-point expansions. They showed that these algorithms are much more robust than it is usually believed: the returned result makes sense even when the rounding function is not round-to-nearest, and they are almost immune to overflow. This work has been submitted [36].

Computing error bounds without changing the rouding mode. S. Boldo has created an algorithm to compute a correct and tight rounding error bound for a floating-point computation. The rounding error can be bounded by folklore formulas, such as $\varepsilon|x|$ or $\varepsilon|\circ(x)|$. This gets more complicated when underflow is taken into account. To compute this error bound in practice, a directed rounding is usually used. This work describes an algorithm that computes a correct bound using only rounding to nearest, therefore without requiring a costly change of the rounding mode. This is formally proved using the Coq formal proof assistant to increase the trust in this algorithm. This has been published at the 9th International Workshop on Numerical Software Verification [17].

Floating-Point Computations and Iterators. S. Boldo has worked on the formal verification of a floating-point case study where the common iterators `fold_left` and `fold_right` have not the wanted behaviors. She then had to define other iterators, which are very similar in most cases, but that do behave well in our case study. This has been published at the 1st Workshop on High-Consequence Control Verification [31].

COMMANDS Project-Team

7. New Results

7.1. Optimal control of ordinary and partial differential equations

7.1.1. On the Design of Optimal Health Insurance Contracts under Ex Post Moral Hazard

Participant: Pierre Martinon.

With Pierre Picard and Anasuya Raj, Ecole Polytechnique.
We analyze in [27] the design of optimal medical insurance under ex post moral hazard, i.e., when illness severity cannot be observed by insurers and policyholders decide on their health expenditures. We characterize the trade-o§ between ex ante risk sharing and ex post incentive compatibility, in an optimal revelation mechanism under hidden information and risk aversion. We establish that the optimal contract provides partial insurance at the margin, with a deductible when insurersí rates are a§ected by a positive loading, and that it may also include an upper limit on coverage. We show that the potential to audit the health state leads to an upper limit on out-of-pocket expenses.

7.1.2. Optimal control of infinite dimensional bilinear systems: application to the heat and wave equations

Participants: J. Frédéric Bonnans, Axel Kröner.

With Soledad Aronna, FGV, Rio de Janeiro. In this paper [13] we consider second order optimality conditions for a bilinear optimal control problem governed by a strongly continuous semigroup operator, the control entering linearly in the cost function. We derive first and second order optimality conditions, taking advantage of the Goh transform. We then apply the results to the heat and wave equations.

7.1.3. Optimal control of PDEs in a complex space setting; application to the Schrödinger equation

Participants: J. Frédéric Bonnans, Axel Kröner.

With Soledad Aronna, FGV, Rio de Janeiro. This paper [22] presents some optimality conditions for abstract optimization problems over complex spaces. We then apply these results to optimal control problems with a semigroup structure. As an application we detail the case when the state equation is the Schrödinger one, with pointwise constraints on the "bilinear'" control. We derive first and second order optimality conditions and address in particular the case that the control enters the state equation and cost function linearly.

7.1.4. Approximation and reduction of optimal control problems in infinite dimension

Participant: Axel Kröner.

With Michael D. Chekroun, UCLA) and H. Liu, Virginia Tech. Nonlinear optimal control problems in infinite dimensions are considered for which we establish approximation theorems and reduction procedures. Approximation theorems and reduction procedures are available in the literature. The originality of our approach relies on a combination of Galerkin approximation techniques with reduction techniques based on finite-horizon parameterizing manifolds. The numerical approximation of the control in a feedback form based on Hamilton-Jacobi-Equation become also affordable within this approach. The approach is applied to optimal control problems of delay differential equations and nonlinear parabolic equations.

7.2. Stochastic control, electricity production and planning

7.2.1. MIDAS: A Mixed Integer Dynamic Approximation Scheme

Participant: J. Frédéric Bonnans.

With Andy Philpott and Faisal Wahid, U. Auckland. Mixed Integer Dynamic Approximation Scheme (MIDAS) [23] is a new sampling-based algorithm for solving finite-horizon stochastic dynamic programs with monotonic Bellman functions. MIDAS approximates these value functions using step functions, leading to stage problems that are mixed integer programs. We provide a general description of MIDAS, and prove its almost-sure convergence to an epsilon-optimal policy when the Bellman functions are known to be continuous, and the sampling process satisfies standard assumptions.

7.2.2. Long term aging : an adaptative weights dynamic programming algorithm

Participants: J. Frédéric Bonnans, Benjamin Heymann, Pierre Martinon.

We introduce [26] a class of optimal control problems with periodic data. A state variable that we call the age of the system represents the negative impact of the operations on the system qualities over time: other things being equal, older systems have higher operating costs. Many industrial problems relate to this class. If we envision to perform an optimization over a large number of periods, there is a tradeoff between minimizing repeatedly the one-period criterion in a short sighted way and taking into account the impact of the decision on the aging speed (which modifies the minimal one period criterion). In general, because the aging process is slow, short term optimization strategies-such as one period sliding horizon strategies-either neglect it or use rule-of-thumb penalization terms in the criterion, which leads to suboptimal solutions. On the other hand, for most applications it is unrealistic to envision a brute-force numerical resolution by dynamic programming of the long term problem because of the computation burden. We introduce a two-scale method to reduce this computation burden. The method relies on Lagrangian duality and some monotony properties. We expose the theoretical foundations of the method and discuss some practical aspects: approximation errors, asymptotic estimation, computation burden, possible extensions, etc. Since our initial motivation was the difficulty to take long term battery aging in Energy Management Systems into account, we implement the method on a toy long term microgrid energy management problem.

7.2.3. Continuous Optimal Control Approaches to Microgrid Energy Management

Participants: J. Frédéric Bonnans, Benjamin Heymann, Pierre Martinon.

With Francisco Silva XLIM, U. Limoges, Fernando Lanas and Guillermo Jimenez, U. Chile.

We propose in [18] a novel method for the microgrid energy management problem by introducing a continuous-time, rolling horizon formulation. The energy management problem is formulated as a deterministic optimal control problem (OCP). We solve (OCP) with two classical approaches: the direct method [1], and Bellman's Dynamic Programming Principle (DPP) [2]. In both cases we use the optimal control toolbox BOCOP [3] for the numerical simulations. For the DPP approach we implement a semi-Lagrangian scheme [4] adapted to handle the optimization of switching times for the on/off modes of the diesel generator. The DPP approach allows for an accurate modeling and is computationally cheap. It finds the global optimum in less than 3 seconds, a CPU time similar to the Mixed Integer Linear Programming (MILP) approach used in [5]. We achieve this performance by introducing a trick based on the Pontryagin Maximum Principle (PMP). The trick increases the computation speed by several orders and also improves the precision of the solution. For validation purposes, simulation are performed using datasets from an actual isolated microgrid located in northern Chile. Results show that DPP method is very well suited for this type of problem when compared with the MILP approach.

7.2.4. A Stochastic Continuous Time Model for Microgrid Energy Management

Participants: J. Frédéric Bonnans, Benjamin Heymann.

With Francisco Silva XLIM U. Limoges, Guillermo Jimenez, U. Chile.

We propose in [20] a novel stochastic control formulation for the microgrid energy management problem and extend previous works on continuous time rolling horizon strategy to uncertain demand. We modelize the demand dynamics with a stochastic differential equation. We decompose this dynamics into three terms: an average drift, a time-dependent mean-reversion term and a Brownian noise. We use BOCOPHJB for the numerical simulations. This optimal control toolbox implements a semi-Lagrangian scheme and handle the

optimization of switching times required for the discrete on/off modes of the diesel generator. The scheme allows for an accurate modelling and is computationally cheap as long as the state dimension is small. As described in previous works, we use a trick to reduce the search of the optimal control values to six points. This increases the computation speed by several orders. We compare this new formulation with the deterministic control approach using data from an isolated microgrid located in northern Chile.

7.2.5. Mechanism Design and Auctions for Electricity Network

Participant: Benjamin Heymann.

With Alejandro Jofré, CMM - Center for Mathematical Modeling, U. Chile, Santiago. We present in [25] some key aspects of wholesale electricity markets modeling and more specifically focus our attention on auctions and mechanism design. Some of the results arising from those models are the computation of an optimal allocation for the Independent System Operator, the study of the equilibria (existence and unicity in particular) and the design of mechanisms to increase the social surplus. From a more general perspective, this field of research provides clues to discuss how wholesale electricity market should be regulated. We start with a general introduction and then present some results the authors obtained recently. We also briefly expose some undergoing related work. As an illustrative example, a section is devoted to the computation of the Independent System Operator response function for a symmetric binodal setting with piece-wise linear production cost functions.

7.2.6. Mechanism design and allocation algorithms for network markets with piece-wise linear costs and externalities

Participant: Benjamin Heymann.

With Alejandro Jofré, CMM - Center for Mathematical Modeling, U. Chile, Santiago. In [24], motivated by market power in electricity market, we introduce a mechanism design for simplified markets of two agents with linear production cost functions. In standard procurement auctions, the market power resulting from the quadratic transmission losses allow the producers to bid above their true value (i.e. production cost). The mechanism proposed in the previous paper reduces the producers margin to the society benefit. We extend those results to a more general market made of a finite number of agents with piecewise linear cost functions, which make the problem more difficult, but at the same time more realistic. We show that the methodology works for a large class of externalities. We also provide two algorithms to solve the principal allocation problem.

7.2.7. Variational analysis for options with stochastic volatility and multiple factors

Participants: J. Frédéric Bonnans, Axel Kröner.

In this ongoing work we discuss the variational analysis for stochastic volatility models with correlation and their applications for the pricing equations for European options is discussed. The considered framework is based on weigthed Sobolev spaces. Furthermore, to verify continuity of the rate term in the pricing equation an approach based on commutator analysis is developed.

DEFI Project-Team

7. New Results

7.1. Methods for inverse problems

7.1.1. Identifying defects in an unknown background using differential measurements

L. Audibert and H. Haddar

In the framework of the PhD thesis of Lorenzo Audibert we studied non destructive testing of concrete using ultrasonic waves, and more generallly imaging in complex heterogeneous media. We assume that measurements are multistatic, which means that we record the scattered field on different points by using several sources. For this type of data we wish to build methods that are able to image the obstacle that created the scattered field. We use qualitative methods in this work, which only provide the support of the object independently from its physical property. The first part of this thesis consists of a theoretical analysis of the Linear Sampling Method. Such analysis is done in the framework of regularization theory, and our main contribution is to provide and analyze a regularization term that ensures good theoretical properties. Among those properties we were able to demonstrate that when the regularization parameter goes to zero, we actually construct a sequence of functions that strongly converges to the solution of the interior transmission problem. This behavior gives a central place to the interior transmission problem as it allows describing the asymptotic solution of our regularized problem. Using this characterization of our solution, we are able to give the optimal reconstruction we can get from our method. More importantly this description of the solution allows us to compare the solution coming from two different datasets. Based on the result of this comparison, we manage to produce an image of the connected component that contains the defect which appears between two measurement campaigns and this regardless of the medium. This method is well suited for the characteristics of the microstructure of concrete as shown on several numerical examples with realistic concrete-like microstructure. Finally, we extend our theoretical results to the case of limited aperture, anisotropic medium and elastic waves, which correspond to the real physics of the ultrasounds

7.1.2. Generalized linear sampling method for elastic-wave sensing of heterogeneous fractures

B. Guzina, H. Haddar and F. Pourahmadian

A theoretical foundation is developed for active seismic reconstruction of fractures endowed with spatially-varying interfacial condition (e.g. partially-closed fractures, hydraulic fractures). The proposed indicator functional carries a superior localization property with no significant sensitivity to the fracture's contact condition, measurement errors, and illumination frequency. This is accomplished through the paradigm of the F-factorization technique and the recently developed Generalized Linear Sampling Method (GLSM) applied to elastodynamics. The direct scattering problem is formulated in the frequency domain where the fracture surface is illuminated by a set of incident plane waves, while monitoring the induced scattered field in the form of (elastic) far-field patterns. The analysis of the well-posedness of the forward problem leads to an admissibility condition on the fracture's (linearized) contact parameters. This in turn contributes toward establishing the applicability of the F-factorization method, and consequently aids the formulation of a convex GLSM cost functional whose minimizer can be computed without iterations. Such minimizer is then used to construct a robust fracture indicator function, whose performance is illustrated through a set of numerical experiments. For completeness, the results of the GLSM reconstruction are compared to those obtained by the classical linear sampling method.

7.1.3. Invisibility in scattering theory

L. Chesnel, A.-S. Bonnet-Ben Dhia and S.A. Nazarov

We are interested in a time harmonic acoustic problem in a waveguide with locally perturbed sound hard walls. We consider a setting where an observer generates incident plane waves at $-\infty$ and probes the resulting scattered field at $-\infty$ and $+\infty$. Practically, this is equivalent to measure the reflection and transmission coefficients respectively denoted R and T. In a recent work, a technique has been proposed to construct waveguides with smooth walls such that $R = 0$ and $|T| = 1$ (non reflection). However the approach fails to ensure $T = 1$ (perfect transmission without phase shift). First we establish a result explaining this observation. More precisely, we prove that for wavenumbers smaller than a given bound $k_{\star}$ depending on the geometry, we cannot have $T = 1$ so that the observer can detect the presence of the defect if he/she is able to measure the phase at $+\infty$. In particular, if the perturbation is smooth and small (in amplitude and in width), $k_{\star}$ is very close to the threshold wavenumber. Then, in a second step, we change the point of view and, for a given wavenumber, working with singular perturbations of the domain, we show how to obtain $T = 1$. In this case, the scattered field is exponentially decaying both at $-\infty$ and $+\infty$. We implement numerically the method to provide examples of such undetectable defects.

7.1.4. Nanoparticles volume determination from SAXS measurements

H. Haddar and Z. Jiang

The aim of this work is to develop a fully automatic method for the reconstruction of the volume distribution of polydisperse non-interacting nanoparticles with identical shapes from Small Angle X-ray Scattering measurements. In the case of diluted systems we proposed a method that solves a maximum likelihood problem with a positivity constraint on the solution by means of an Expectation Maximization iterative scheme coupled with a robust stopping criterion. We prove that the stopping rule provides a regularization method according to an innovative notion of regularization specifically defined for inverse problems with Poisson data. Such a regularization, together with the positivity constraint results in high fidelity quantitative reconstructions of particle volume distributions making the method particularly effective in real applications. We tested the performance of the method on synthetic data in the case of uni- and bi-modal particle volume distributions. We extended the method to the case of dense solutions where the inverse problem becomes non linear. A specific fix-point algorithm has been proposed and convergence has been tested against synthetic data. The developement of this research topic is ongoing under the framework of Saxsize.

7.1.5. Identifying defects in unknown periodic layers

H. Haddar and T.P. Nguyen

We investigate the inverse problem where one is interested in reconstructing the support of a perturbation in a periodic media from measurements of scattered waves. We are concerned with the design of a sampling method that would reconstruct the support of inhomogeneities without reconstructing the index of refraction. The development of sampling methods has gained a large interest in recent years and many methods have been introduced in the literature to deal with a variety of problems and we refer to [1] for an account of recent developments of these methods. Up to our knowledge, the sampling methods for locally perturbed infinite periodic layers has not been treated in the literature. Even thought this problem is the one that motivates our study, we considered a slightly different problem that will be referred to as the ML-periodic problem: it corresponds with a locally perturbed infinite periodic layer with period L that has been reduced to a domain of size ML (with M a sufficiently large parameter) with periodic boundary conditions. This is mainly for technical reasons since our analysis for the newly introduced differential imaging functional heavily rely on the discrete Floquet-Bloch transform.

The main contribution of our work is the design of a new sampling method that enable the imaging of the defect location without reconstructing the L periodic background. This method is in the spirit of the Differential LSM introduced above for the imaging of defects in complex backgrounds using differential measurements. However, in the present case we propose a method that does not require the measurement operator for the background media. We exploit the L periodicity of the background and the Floquet-Bloch transform to design a differential criterion between different periods. This criterion is based on the study of sampling methods for the ML-periodic media where a single Floquet-Bloch mode is used. This study constitutes the main theoretical ingredient for our method. The sampling operator for a single Floquet-Bloch mode somehow plays the role of

the measurement operator for the background media. Indeed the main interest for this new sampling method is that it is capable of identifying the defect even thought classical sampling methods fail in obtaining high fidelity reconstructions of the (complex) background media.

7.1.6. Identification of small objects with near-field data in quasi-backscattering configurations

H. Haddar and M. Lakhal

We present a new sampling method for detecting targets (small inclusions or defects) immersed in a homogeneous medium in three-dimensional space, from measurements of acoustic scattered fields created by point source incident waves. We consider the harmonic regime and a data setting that corresponds with quasi-backscattering configuration: the data is collected by a set a receivers that are distributed on a segment centered at the source position and the device is swept along a path orthogonal to the receiver line. We assume that the aperture of the receivers is small compared with the distance to the targets. Considering the asymptotic form of the scattered field as the size of the targets goes to zero and the small aperture approximation, one is able to derive a special expression for the scattered field. In this expression a separation of the dependence of scattered field on the source location and the distance source-target is performed. This allows us to propose a sampling procedure that characterizes the targets location in terms of the range of a near-field operator constructed from available data. Our procedure is similar to the one proposed by Haddar-Rezac for far-field configurations. The reconstruction algorithm is based on the MUSIC (Multiple SIgnal Classification) algorithm.

7.1.7. Nondestructive testing of the delaminated interface between two materials

F. Cakoni, I. De Teresa, H. Haddar and P. Monk

We consider the problem of detecting if two materials that should be in contact have separated or delaminated. The goal is to find an acoustic technique to detect the delamination. We model the delamination as a thin opening between two materials of different acoustic properties, and using asymptotic techniques we derive a asymptotic model where the delaminated region is replaced by jump conditions on the acoustic field and flux. The asymptotic model has potential singularities due to the edges of the delaminated region, and we show that the forward problem is well posed for a large class of possible delaminations. We then design a special Linear Sampling Method (LSM) for detecting the shape of the delamination assuming that the background, undamaged, state is known. Finally we show, by numerical experiments, that our LSM can indeed determine the shape of delaminated regions.

7.2. Shape and topology optimization

7.2.1. Second-order shape derivatives along normal trajectories, governed by Hamilton-Jacobi equations

G. Allaire, E. Cancès and J.-L. Vié

In this work we introduce a new variant of shape differentiation which is adapted to the deformation of shapes along their normal direction. This is typically the case in the level-set method for shape optimization where the shape evolves with a normal velocity. As all other variants of the orginal Hadamard method of shape differentiation, our approach yields the same first order derivative. However, the Hessian or second-order derivative is different and somehow simpler since only normal movements are allowed. The applications of this new Hessian formula are twofold. First, it leads to a novel extension method for the normal velocity, used in the Hamilton-Jacobi equation of front propagation. Second, as could be expected, it is at the basis of a Newton optimization algorithm which is conceptually simpler since no tangential displacements have to be considered. Numerical examples are given to illustrate the potentiality of these two applications. The key technical tool for our approach is the method of bicharacteristics for solving Hamilton-Jacobi equations. Our new idea is to differentiate the shape along these bicharacteristics (a system of two ordinary differential equations).

7.2.2. Introducing a level-set based shape and topology optimization method for the wear of composite materials with geometric constraints

G. Allaire, F. Feppon, G. Michailidis, M.S. Sidebottom, B.A. Krick and N. Vermaak

The wear of materials continues to be a limiting factor in the lifetime and performance of mechanical systems with sliding surfaces. As the demand for low wear materials grows so does the need for models and methods to systematically optimize tribological systems. Elastic foundation models offer a simplified framework to study the wear of multimaterial composites subject to abrasive sliding. Previously, the evolving wear profile has been shown to converge to a steady-state that is characterized by a time-independent elliptic equation. In this article, the steady-state formulation is generalized and integrated with shape optimization to improve the wear performance of bi-material composites. Both macroscopic structures and periodic material microstructures are considered. Several common tribological objectives for systems undergoing wear are identified and mathematically formalized with shape derivatives. These include (i) achieving a planar wear surface from multimaterial composites and (ii) minimizing the run-in volume of material lost before steady-state wear is achieved. A level-set based topology optimization algorithm that incorporates a novel constraint on the level-set function is presented. In particular, a new scheme is developed to update material interfaces; the scheme (i) conveniently enforces volume constraints at each iteration, (ii) controls the complexity of design features using perimeter penalization, and (iii) nucleates holes or inclusions with the topological gradient. The broad applicability of the proposed formulation for problems beyond wear is discussed, especially for problems where convenient control of the complexity of geometric features is desired.

7.2.3. Geometric constraints for shape and topology optimization in architectural design

G. Allaire, C. Dapogny, A. Faure, G. Michailidis, A. Couvelas and R. Estevez

This work proposes a shape and topology optimization framework oriented towards conceptual architectural design. A particular emphasis is put on the possibility for the user to interfere on the optimization process by supplying information about his personal taste. More precisely, we formulate three novel constraints on the geometry of shapes; while the first two are mainly related to aesthetics, the third one may also be used to handle several fabrication issues that are of special interest in the device of civil structures. The common mathematical ingredient to all three models is the signed distance function to a domain, and its sensitivity analysis with respect to perturbations of this domain; in the present work, this material is extended to the case where the ambient space is equipped with an anisotropic metric tensor. Numerical examples are discussed in two and three space dimensions.

7.2.4. Modal basis approaches in shape and topology optimization of frequency response problems

G. Allaire and G. Michailidis

The optimal design of mechanical structures subject to periodic excitations within a large frequency interval is quite challenging. In order to avoid bad performances for non-discretized frequencies, it is necessary to finely discretize the frequency interval, leading to a very large number of state equations. Then, if a standard adjoint-based approach is used for optimization, the computational cost (both in terms of CPU and memory storage) may be prohibitive for large problems, especially in three space dimensions. The goal of the present work is to introduce two new non-adjoint approaches for dealing with frequency response problems in shape and topology optimization. In both cases, we rely on a classical modal basis approach to compute the states, solutions of the direct problems. In the first method, we do not use any adjoint but rather directly compute the shape derivatives of the eigenmodes in the modal basis. In the second method, we compute the adjoints of the standard approach by using again the modal basis. The numerical cost of these two new strategies are much smaller than the usual ones if the number of modes in the modal basis is much smaller than the number of discretized excitation frequencies. We present numerical examples for the minimization of the dynamic compliance in two and three space dimensions.

7.3. Direct scattering problems

7.3.1. Finite element methods for eigenvalue problems with sign-changing coefficients

C. Carvalho, P. Ciarlet and L. Chesnel

We consider a class of eigenvalue problems involving coefficients changing sign on the domain of interest. We analyse the main spectral properties of these problems according to the features of the coefficients. Under some assumptions on the mesh, we study how one can use classical finite element methods to approximate the spectrum as well as the eigenfunctions while avoiding spurious modes. We also prove localisation results of the eigenfunctions for certain sets of coefficients.

7.3.2. A Volume integral method for solving scattering problems from locally perturbed periodic layers

H. Haddar and T.P. Nguyen

We investigate the scattering problem for the case of locally perturbed periodic layers in R^d, $d = 2, 3$. Using the Floquet-Bloch transform in the periodicity direction we reformulate this scattering problem as an equivalent system of coupled volume integral equations. We then apply a spectral method to discretize the obtained system after periodization in the direction orthogonal to the periodicity directions of the medium. The convergence of this method is established and validating numerical results are provided.

7.4. Asymptotic Analysis

7.4.1. Small obstacle asymptotics for a non linear problem

L. Chesnel, X. Claeys and S.A. Nazarov

We study a 2D semi-linear equation in a domain with a small Dirichlet obstacle of size δ. Using the method of matched asymptotic expansions, we compute an asymptotic expansion of the solution as δ tends to zero. Its relevance is justified by proving a rigorous error estimate. We also construct an approximate model, based on an equation set in the limit domain without the small obstacle, which provides a good approximation of the far field of the solution of the original problem. The interest of this approximate model lies in the fact that it leads to a variational formulation which is very simple to discretize. We present numerical experiments to illustrate the analysis.

7.4.2. Influence of the geometry on plasmonic waves

L. Chesnel X. Claeys and S.A. Nazarov

In the modeling of plasmonic technologies in time harmonic regime, one is led to study the eigenvalue problem $-\div(\sigma\nabla u) = \lambda u$ (P), where σ is a physical coefficient positive in some region Ω_+ and negative in some other region Ω_-. We highlight an unusual instability phenomenon for the source term problem associated with (P): for certain configurations, when the interface between Ω_+ and Ω_- presents a rounded corner, the solution may depend critically on the value of the rounding parameter. We explain this property studying the eigenvalue problem (P). We provide an asymptotic expansion of the eigenvalues and prove error estimates. We establish an oscillatory behaviour of the eigenvalues as the rounding parameter of the corner tends to zero. These theoretical results are illustrated by numerical experiments.

7.4.3. Instability of dielectrics and conductors in electrostatic fields

G. Allaire and J. Rauch

This work proves most of the assertions in section 116 of Maxwell's treatise on electromagnetism. The results go under the name Earnshaw's Theorem and assert the absence of stable equilibrium configurations of conductors and dielectrics in an external electrostatic field.

7.4.4. Optimization of dispersive coefficients in the homogenization of the wave equation in periodic structures

G. Allaire and T. Yamada

We study dispersive effects of wave propagation in periodic media, which can be modelled by adding a fourth-order term in the homogenized equation. The corresponding fourth-order dispersive tensor is called Burnett tensor and we numerically optimize its values in order to minimize or maximize dispersion. More precisely, we consider the case of a two-phase composite medium with an 8-fold symmetry assumption of the periodicity cell in two space dimensions. We obtain upper and lower bound for the dispersive properties, along with optimal microgeometries.

7.4.5. Homogenization of Stokes System using Bloch Waves

G. Allaire, T. Ghosh and M. Vanninathan

In this work, we study the Bloch wave homogenization for the Stokes system with periodic viscosity coefficient. In particular, we obtain the spectral interpretation of the homogenized tensor. The presence of the incompressibility constraint in the model raises new issues linking the homogenized tensor and the Bloch spectral data. The main difficulty is a lack of smoothness for the bottom of the Bloch spectrum, a phenomenon which is not present in the case of the elasticity system. This issue is solved in the present work, completing the homogenization process of the Stokes system via the Bloch wave method.

7.5. Diffusion MRI

7.5.1. Adapting the Kärger model to account for finite diffusion-encoding pulses in diffusion MRI

H. Haddar, J.R. Li and S. Schiavi

Diffusion magnetic resonance imaging (dMRI) is an imaging modality that probes the diffusion characteristics of a sample via the application of magnetic field gradient pulses. If the imaging voxel can be divided into different Gaussian diffusion compartments with inter-compartment exchange governed by linear kinetics, then the dMRI signal can be described by the Kärger model, which is a well-known model in NMR. However, the Kärger model is limited to the case when the duration of the diffusion-encoding gradient pulses is short compared to the time delay between the start of the pulses. Under this assumption, the time at which to evaluate the Kärger model to obtain the dMRI signal is unambiguously the delay between the pulses. Recently, a new model of the dMRI signal, the Finite-Pulse Kärger (FPK) model, was derived for arbitrary diffusion gradient profiles. Relying on the FPK model, we show that when the duration of the gradient pulses is not short, the time at which to evaluate the Kärger model should be the time delay between the start of the pulses, shortened by one third of the pulse duration. With this choice, we show the sixth order convergence of the Kärger model to the FPK model in the non-dimensionalized pulse duration.

7.5.2. A macroscopic model for the diffusion MRI signal accounting for time-dependent diffusivity

H. Haddar, J.R. Li and S. Schiavi

An important quantity measured in dMRI in each voxel is the Apparent Diffusion Coefficient (ADC) and it is well-established from imaging experiments that, in the brain, *in-vivo*, the ADC is dependent on the measured diffusion time. To aid in the understanding and interpretation of the ADC, using homogenization techniques, we derived a new asymptotic model for the dMRI signal from the Bloch-Torrey equation governing the water proton magnetization under the influence of diffusion-encoding magnetic gradient pulses. Our new model was obtained using a particular choice of scaling for the time, the biological cell membrane permeability, the diffusion-encoding magnetic field gradient strength, and a periodicity length of the cellular geometry. The ADC of the resulting model is dependent on the diffusion time. We numerically validated this model for a wide range of diffusion times for two dimensional geometrical configurations.

7.5.3. Quantitative DLA-based Compressed Sensing for MEMRI Acquisitions

P. Svehla, K.-V. Nguyen, J.-R. Li and L. Ciobanu

High resolution Manganese Enhanced Magnetic Resonance Imaging (MEMRI) has great potential for functional imaging of live neuronal tissue at single neuron scale. However, reaching high resolutions often requires long acquisition times which can lead to reduced image quality due to sample deterioration and hardware instability. Compressed Sensing (CS) techniques offer the opportunity to significantly reduce the imaging time. The purpose of this work is to test the feasibility of CS acquisitions based on Diffusion Limited Aggregation (DLA) sampling patterns for high resolution quantitative MEMRI imaging. Fully encoded and DLA-CS MEMRI images of Aplysia californica neural tissue were acquired on a 17.2T MRI system. The MR signal corresponding to single, identified neurons was quantified for both versions of the T1 weighted images. Results: For a 50% undersampling, DLA-CS leads to signal intensity differences, measured in individual neurons, of approximately 1.37% when compared to the fully encoded acquisition, with minimal impact on image spatial resolution. At the undersampling ratio of 50%, DLA-CS is capable of accurately quantifying signal intensities in MEMRI acquisitions. Depending on the image signal to noise ratio, higher undersampling ratios can be used to further reduce the acquisition time in MEMRI based functional studies of living tissues.

7.5.4. The time-dependent diffusivity in the abdominal ganglion of Aplysia californica, comparing experiments and simulations

K.-V. Nguyen, D. Le Bihan, L. Ciobanu and J.-R. Li

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 shed light on this relationship for more complex organisms. We measured the dMRI signal at several diffusion times in the abdominal ganglion and performed simulations of water diffusion in geometries obtained after segmenting high resolution T2-weighted images and incorporating known information about the cellular structure from the literature. By fitting the experimental signal to the simulated signal for several types of cells in the abdominal ganglion at a wide range of diffusion times, we obtained estimates of the intrinsic diffusion coefficient in the nucleus and the cytoplasm. We also evaluated the reliability of using an existing formula for the time-dependent diffusion coefficient to estimate cell size.

7.5.5. A two pool model to describe the IVIM cerebral perfusion

G. Fournet, J.-R. Li, A.M. Cerjanic, B.P. Sutton, L. Ciobanu and D. Le Bihan

IntraVoxel Incoherent Motion (IVIM) is a magnetic resonance imaging (MRI) technique capable of measuring perfusion-related parameters. In this manuscript, we show that the mono-exponential model commonly used to process IVIM data might be challenged, especially at short diffusion times. Eleven rat datasets were acquired at 7T using a diffusion-weighted pulsed gradient spin echo sequence with b-values ranging from 7 to 2500 s/mm^2 at 3 diffusion times. The IVIM signals, obtained by removing the diffusion component from the raw MR signal, were fitted to the standard mono-exponential model, a bi-exponential model and the Kennan model. The Akaike information criterion used to find the best model to fit the data demonstrates that, at short diffusion times, the bi-exponential IVIM model is most appropriate. The results obtained by comparing the experimental data to a dictionary of numerical simulations of the IVIM signal in microvascular networks support the hypothesis that such a bi-exponential behavior can be explained by considering the contribution of two vascular pools: capillaries and somewhat larger vessels.

7.5.6. The influence of acquisition parameters on the metrics of the bi-exponential IVIM model

G. Fournet, J.-R. Li, D. Le Bihan and L. Ciobanu

The IntraVoxel Incoherent Motion (IVIM) MRI signal, typically described as a mono-exponential decay, can sometimes be better modeled as a bi-exponential function accounting for two vascular pools, capillaries and medium-size vessels. The goal of this work is to define precisely in which conditions the IVIM signal shape becomes bi-exponential and to understand the evolution of the IVIM outputs with different acquisition parameters. Rats were scanned at 7T and 11.7T using diffusion-weighted pulsed-gradient spin-echo (SE)

and stimulated-echo (STE) sequences with different repetition times (TR) and diffusion encoding times. The obtained IVIM signals were fit to the mono- and bi-exponential models and the output parameters compared. The bi-exponential and mono-exponential models converge at long diffusion encoding times and long TRs. The STE is less sensitive to inflow effects present at short TRs, leading to a smaller volume fraction for the fast pool when compared to the SE sequence. The two vascular components are more easily separated at short diffusion encoding times, short TRs and when using a SE sequence. The volume fractions of the two blood pools depend on the pulse sequence, TR and diffusion encoding times while the pseudo-diffusion coefficients are only affected by the diffusion encoding time.

DISCO Project-Team

5. New Results

5.1. Characterizing the Codimension of Zero Singularities for Time-Delay Systems: A Link with Vandermonde and Birkhoff Incidence Matrices

Participants: Islam Boussaada, Silviu-Iulian Niculescu.

The analysis of time-delay systems mainly relies on detecting and understanding the spectral values bifurcations when crossing the imaginary axis. We have dealed with the zero singularity, essentially when the zero spectral value is multiple. The simplest case in such a configuration is characterized by an algebraic multiplicity two and a geometric multiplicity one, known as the Bogdanov-Takens singularity. Moreover, in some cases the codimension of the zero spectral value exceeds the number of the coupled scalar-differential equations. Nevertheless, to the best of the author's knowledge, the bounds of such a multiplicity have not been deeply investigated in the literature. It is worth mentioning that the knowledge of such an information is crucial for nonlinear analysis purposes since the dimension of the projected state on the center manifold is none other than the sum of the dimensions of the generalized eigenspaces associated with spectral values with zero real parts. Motivated by a control-oriented problems, we have provided an answer to this question for time-delay systems, taking into account the parameters' algebraic constraints that may occur in applications. We emphasize the link between such a problem and the incidence matrices associated with the Birkhoff interpolation problem. In this context, symbolic algorithms for LU-factorization for functional confluent Vandermonde as well as some classes of bivariate functional Birkhoff matrices are also proposed [11].

5.2. Tracking the Algebraic Multiplicity of Crossing Imaginary Roots for Generic Quasipolynomials: A Vandermonde-Based Approach

Participants: Islam Boussaada, Silviu-Iulian Niculescu.

A standard approach in analyzing dynamical systems consists in identifying and understanding the eigenvalues bifurcations when crossing the imaginary axis. Efficient methods for crossing imaginary roots identification exist. However, to the best of the author's knowledge, the multiplicity of such roots was not deeply investigated. We have emphasized [12] that the multiplicity of the zero spectral value can exceed the number of the coupled scalar delay-differential equations and a constructive approach Vandermonde-based allowing to an adaptive bound for such a multiplicity is provided. Namely, it is shown that the zero spectral value multiplicity depends on the system structure (number of delays and number of non zero coefficients of the associated quasipolynomial) rather than the degree of the associated quasipolynomial. We have extended the constructive approach in investigating the multiplicity of crossing imaginary roots $j\omega$ where $\omega \neq 0$ and establishes a link with a new class of functional confluent Vandermonde matrices. A symbolic algorithm for computing the LU-factorization for such matrices is provided. As a byproduct of the proposed approach, a bound sharper than the Polya-Szegö generic bound arising from the principle argument is established.

5.3. Coprimeness of fractional representations

Participants: Catherine Bonnet, Le Ha Vy Nguyen, Yutaka Yamamoto [Kyoto Univ].

Coprimeness of a fractional representation plays various crucial roles in many different contexts, for example, stabilization of a given plant, minimality of a state space representation, etc. It should be noted however that coprimeness depends crucially on the choice of a ring (or algebra) where such a representation is taken, which reflects the choice of a plant, and particular problems that one studies. Such relationships are particularly delicate and interesting when dealing with infinite-dimensional systems. We have disucssed various coprimeness issues for different rings, typically for H_∞ and pseudorational transfer functions. The former is related to H_∞-stabilizability, and the latter to controllability of behaviors. We have also given some intricate examples where a seemingly non-coprime factorization indeed turns out to be a coprime factorization over H_∞ [28], [29].

5.4. Output-feddback control design for time-delay systems

Participants: Catherine Bonnet, Caetano Cardeliquio, Matheus Souza [FEEC-UNICAMP], André Fioravanti [FEM-UNICAMP].

We presented new results on H_∞-control synthesis, via output-feedback, for time-delay linear systems. We extend the use of a finite order LTI system, called comparison system, to design a controller which depends not only on the output at the present time and maximum delay, but also on an arbitrary number of values between those. This approach allows us to increase the maximum stable delay without requiring any additional information.

5.5. Backstepping control design through the introduction of delays

Participants: Frederic Mazenc, Michael Malisoff [LSU], Jerome Weston [LSU].

We provided new backstepping results for time-varying systems with input delays. The results were obtained by the introduction of constant 'artificial' pointwise delays in the input. Thus they are significantly different from backstepping results for systems with delay in the input as presented in previous contributions.

I) The novelty of the contribution in [18] is in the bounds on the controls, and the facts that (i) one does not need to compute any Lie derivatives to apply the proposed controls, (ii) the controls have no distributed terms, and (iii) no differentiability conditions on the available controls for the subsystems are needed. The latter aspect is of paramount importance from an applied point of view.

II) In [43], we extended [18]. We provided new globally stabilizing backstepping controls for single input systems in a partially linear form. Instead of measuring the full state, the feedbacks use current and several time lagged values of a function of the state of the nonlinear subsystem (and have no distributed terms). We also allowed input delays. This improves on [18], since we allowed an arbitrary number of integrators whereas [18] is limited to one integrator.

5.6. Switched Nonlinear Systems

Participants: Frederic Mazenc, Yue-E Wang [Shaanxi Normal University], Xi-Ming Sun [Dalian University of Technology].

We considered in [26] a class of nonlinear time-varying switched control systems for which stabilizing feedbacks are available. We analyzed the effect of the presence of a delay in the input of switched nonlinear systems with an external disturbance. By contrast with most of the contributions available in the literature, we did not assume that all the subsystems of the switched system we consider are stable when the delay is present. Through a Lyapunov approach, we derived sufficient conditions in terms of size of the delay, ensuring the global exponential stability of the switched system. Moreover, under appropriate conditions, the input-to-state stability of the system with respect to an external disturbance was established.

5.7. Studies of systems with long delays

Participants: Frederic Mazenc, Michael Malisoff [LSU], Emilia Fridman [Tel-Aviv University].

We solved several problems of observer and control designs pertaining to the fundamental (and difficult) case where a delay in the input is too long for being neglected.

I) We considered in [17] the problem of stabilizing a linear continuous-time system with discrete-time measurements and a sampled input with a pointwise constant delay. In a first part, we designed a continuous-discrete observer which converges when the maximum time interval between two consecutive measurements is sufficiently small. In a second part, we constructed a dynamic output feedback by using a technique which is strongly reminiscent of the so called 'reduction model approach'. It stabilizes the system when the maximal time between two consecutive sampling instants is sufficiently small. No limitation on the size of the delay was imposed and an ISS property with respect to additive disturbances was established.

II) We solved stabilization problems for linear time-varying systems under input delays. We showed how changes of coordinates lead to systems with time invariant drifts, which are covered by the reduction model method and which lead to the problem of stabilizing a time-varying system without delay. For continuous-time periodic systems, we used Floquet's theory to find the changes of coordinates. We also proved an analogue for discrete time systems, through an original discrete-time extension of Floquet's theory [19].

III) In [21] and [42], we proposed a prediction based stabilization approach for a general class of nonlinear time-varying systems with pointwise delay in the input. It is based on a recent new prediction strategy, which makes it possible to circumvent the problem of constructing and estimating distributed terms in the expression for the stabilizing control laws. We observed that our result applies in cases where other recent results do not, including notably the case where a time-varying delay is present.

5.8. Extension of the Razumikhin's theorem

Participants: Frederic Mazenc, Michael Malisoff [LSU].

The Razumikhin's Theorem is a major extension of the Lypunov function theory, making possible to establish the global asymptotic stability of nonlinear systems with delays. It is especially efficient when the delays are time-varying. We provide in [41] an extension of this theorem for continuous-time time-varying systems with time-varying delays. Our result uses a novel 'strictification' technique for converting a nonstrict Lyapunov function into a strict one. Our examples show how our method can sometimes allow broader classes of allowable delays than the results in the literature.

5.9. Observer design for nonlinear systems

Participant: Ali Zemouche.

A new high-gain observer design method with lower gain compared to the standard high-gain observer was proposed. This new observer, called "HG/LMI" observer is obtained by combining the standard high-gain methodology with the LPVLMI-based technique. Through analytical developments, it is shown how the new observer provides a lower gain. A numerical example was used to illustrate the performance of the new "HG/LMI" observer. The aim of this research is the application of this new observer design to estimate some vehicle variables in autonomous vehicle applications.

5.10. Set invariance for discrete-time delay systems

Participants: Sorin Olaru, Mohammed Laraba [L2S], Silviu Niculescu, Franco Blanchini [Univ. Udine, Italy], Stefano Miani [Univ. Udine, Italy].

The existence of positively invariant sets for linear delay-difference equations was pursued in [15]. We made a survey effort and presented in a unified framework all known necessary and/or sufficient conditions for the existence of invariant sets with respect to dynamical systems described by linear discrete time-delay difference equations (dDDEs). Secondly, we address the construction of invariant sets in the original state space (also called D-invariant sets) by exploiting the forward mappings. The notion of D-invariance is appealing since it provides a region of attraction, which is difficult to obtain for delay systems without taking into account the delayed states in some appropriate extended state space model. The paper contains a sufficient condition for the existence of ellipsoidal D-contractive sets for dDDEs, and a necessary and sufficient condition for the existence of D-invariant sets in relation to linear time-varying dDDE stability. Another contribution is the clarification of the relationship between convexity (convex hull operation) and D-invariance of linear dDDEs. In short, it is shown that the convex hull of the union of two or more D-invariant sets is not necessarily D-invariant, while the convex hull of a non-convex D-invariant set is D-invariant. Positive invariance is an essential concept in control theory, with applications to constrained dynam-ical systems analysis, uncertainty handling as well as related control design problems. It serves as a basic tool in many topics, such as model predictive control, fault tolerant control and reference governor design. Furthermore, there exists a close link between classical stability theory and positive invariant sets. It is worth mentioning that, in Lyapunov theory, invariance is implicitly described.

5.11. Interpolation-based design for constrained dynamical systems

Participants: Sorin Olaru, Nam Nguyen [IFP, France], Per Olof Gutman [Technion, Israel].

A technique is presented in [49] leading to an explicit state feedback solution to the regulation problem for uncertain and/or time-varying linear discrete-time systems with state and control constraints. A piecewise affine control law is provided which not only guarantees recursive feasibility and robust asymptotic stability, but is also optimal for a region of the state space containing the origin.

5.12. Inverse optimality results for constrained control

Participants: Sorin Olaru, Ngoc Anh Nguyen [L2S], Pedro Rodriguez [L2S], Morten Hovd [NTNU Trondheim, Norway], Ioan Necoara [Univ. Politehnica Bucharest, Romania].

Parametric convex programming has received a lot of attention, since it has many applications in chemical engineering, control engineering, signal processing , etc. Further, inverse optimality plays an important role in many contexts, e.g., image processing, motion planning. In this context we introduced [10] a constructive solution of the inverse optimality problem for the class of continuous piecewise affine functions. The main idea is based on the convex lifting concept. Accordingly, an algorithm to construct convex liftings of a given convexly liftable partition have been put forward. Continuous piecewise affine function defined over a polytopic partition of the state space are known to be obtained as the solution of a parametric linear/quadratic programming problem. Regarding linear model predictive control, is shown that any continuous piecewise affine control law can be obtained via a linear optimal control problem with the control horizon at most equal to 2 prediction steps.

5.13. Robustness and fragility of Piecewise affine control laws

Participants: Sorin Olaru, Ngoc Anh Nguyen [L2S], Pedro Rodriguez [L2S], Morten Hovd [NTNU Trondheim, Norway], Georges Bitsoris [Univ. Patras, Greece].

We focus in [9]on the robustness and fragility problem for piecewise affine (PWA) control laws for discrete-time linear system dynamics in the presence of parametric uncertainty of the state space model. A generic geometrical approach will be used to obtain robustness/fragility margins with respect to the positive invariance properties. For PWA control laws defined over a bounded region in the state space, it is shown that these margins can be described in terms of polyhedral sets in parameter space. The methodology is further extended to the fragility problem with respect to the partition defining the controller. Finally, several computational aspects are presented regarding the transformation from the theoretical formulations to explicit representations (vertex/halfspace representation of polytopes) of these sets.

5.14. Distributed robust model predictive control

Participants: Sorin Olaru, Alexandra Grancharova [Technical University of Sofia, Bulgaria].

We presented in a suboptimal approach to distributed closed-loop min-max MPC for uncertain systems consisting of polytopic subsystems with coupled dynamics subject to both state and input constraints. The approach applies the dynamic dual decomposition method and reformulates the original centralized min-max MPC problem into a distributed optimization problem. The suggested approach was illustrated on a simulation example of an uncertain system consisting of two interconnected polytopic subsystems.

5.15. Algebraic Analysis Approach to Linear Functional Systems

Participants: Guillaume Sandou, Mohamed Lotfi Derouiche [Ecole nationale d'Ingénieurs de Tunis], Soufiene Bouallegue [Ecole nationale d'Ingénieurs de Tunis], Joseph Haggège Derouiche [Ecole nationale d'Ingénieurs de Tunis].

In this study, a new Model Predictive Controller (MPC) parameters tuning strategy is proposed using a LabVIEW-based perturbed Particle Swarm Optimization (pPSO). This original LabVIEW implementation of this metaheuristic algorithm is firstly validated on some test functions in order to show its efficiency and validity. The optimization results are compared with the standard PSO approach. The parameters tuning problem, i.e. the weighting factors on the output error and input increments of the MPC algorithm, is then formulated and systematically solved, using the proposed LabVIEW pPSA algorithm. The case of a Magnetic Levitation (MAGLEV) system is investigated to illustrate the robustness and superiority of the proposed pPSO-based tuning MPC approach. All obtained simulation results, as well as the statistical analysis tests for the formulated control problem with and without constraints, are discussed and compared with the Genetic Algorithm Optimization (GAO)-based technique in order to improve the effectiveness of the proposed pPSA-based MPC tuning methodology derouiche:hal-01347041.

5.16. Attitude dynamics, control and observation

Participants: Frederic Mazenc, Maruthi Akella [Univ of Texas], Sunpil Yang [Univ of Texas].

In [27], we addressed the rigid-body attitude tracking problem in the absence of angular velocity measurements. To achieve proportional-derivative feedback control, an angular velocity observer with global exponential convergence was designed based on the Immersion and Invariance (I&I) method. A dynamic scaling factor was introduced to circumvent the integrability condition typically arising in I&I design. Unlike the existing I&I observer for this problem, the estimated angular velocity is defined using a rotation matrix of the current quaternion state to avoid use of an additional filter for the angular velocity estimate. As a result, stability analysis became less complex and the observer structure was further simplified by efficient handling of the Coriolis effect in the observer error dynamics. In the case where proportional-derivative control is combined with the observer, asymptotic convergence of tracking errors was proved while establishing a separation property. Numerical simulations were provided to demonstrate the performance of the proposed observer and the output feedback controller.

5.17. Estimation for vehicle application

Participants: Ali Zemouche, Rajesh Rajamani [University of Minneapolis, USA], Gridsada Phanomchoeng [Chulalongkorn University, Thailand].

A new LMI (Linear Matrix Inequality) design technique is developed to address the problem of circle criterion based $\mathcal{H}_\infty$ observer design for nonlinear systems. The developed technique applies to both locally Lipschitz as well as monotonic nonlinear systems, and allows for nonlinear functions in both the process dynamics and output equations. The LMI design condition obtained is less conservative than all previous results proposed in literature for these classes of nonlinear systems. By judicious use of a modified Young's relation, additional degrees of freedom are included in the observer design. These additional decision variables enable improvements in the feasibility of the obtained LMI. Several recent results in literature are shown to be particular cases of the more general observer design methodology developed in this paper. Illustrative examples are used to show the effectiveness of the proposed methodology. The application of the method to slip angle estimation in automotive applications is discussed and experimental results are presented. This application was the main motivation of this work.

5.18. Observer-based stabilization for lateral vehicle control

Participants: Ali Zemouche, Rajesh Rajamani [University of Minneapolis, USA], Yan Wang [University of Minneapolis, USA].

Recently, motivated by autonomous vehicle control problem, a robust observer based estimated state feedback control design method for an uncertain dynamical system that can be represented as a LTI system connected with an IQC-type nonlinear uncertainty was developed. Different from existing design methodologies in which a convex semidefinite constraint is obtained at the cost of conservatism and unrealistic assumptions, the design of the robust observer state feedback controller is formulated in this paper as a feasibility problem of a bilinear matrix inequality (BMI) constraint. Unfortunately, the search for a feasible solution of a BMI constraint is a NP hard problem in general. The applicability of the linearization method, such as variable change method or congruence transformation, depends on the specific structure of the problem at hand and cannot be generalized. A new sequential LMI optimization method to search for a feasible solution was established. A vehicle lateral control problem was presented to demonstrate the applicability of the proposed algorithm to a real-world estimated state feedback control design.

5.19. Unified model for low-cost high-performance AC drives: the equivalent flux concept

Participants: Guillaume Sandou, Mohamad Koteich [Renault], Abdelmalek Maloum [Renault], Gilles Duc [CentraleSupélec].

This study presents a unified modeling approach of alternating current (AC) machines for low-cost high-performance drives. The Equivalent Flux concept is introduced. Using this concept, all AC machines can be seen as a non-salient synchronous machine with modified (equivalent) rotor flux. Therefore, complex salient-rotor machines models are simplified, and unified shaft-sensorless AC drives can be sought. For this purpose, a unified observer-based structure for rotor-flux position and speed estimation is proposed. The equivalent flux concept generalizes the existing concepts, such as the extended back-electromotive force, the fictitious flux and the active flux.

5.20. Supervision and rescheduling of a mixed CBTC traffic on a suburban railway line

Participants: Guillaume Sandou, Juliette Pochet [SNCF], Sylvain Baro [SNCF].

Railway companies need to achieve higher capacities on existing infrastructures such as high density suburban mainlines. Communication based train control (CBTC) systems have been widely deployed on dedicated subway lines. However, deployment on shared rail infrastructure, where CBTC and non-CBTC trains run, leads to a mixed positioning and controlling system with different precision levels and restrictions. New performance and complexity issues are to arise. In this study, a method for rescheduling adapted to a CBTC system running in a mixed traffic, is introduced. The proposed method is based on a model predictive control (MPC) approach. In each step, an enhanced genetic algorithm with new mutation mechanisms solves the problem to optimize the cost function. It determines the dwell times and running times of CBTC trains, taking into account the non-CBTC trains planning and fixed-block localization. In addition, reordering can be allowed by modifying the problem constraints. The work is supported by a simulation tool developed by SNCF and adapted to mixed traffic study. The approach is illustrated with a case study based on a part of an East/West line in the Paris region network, proving the ability of the method to find good feasible solutions when delays occur in traffic [46].

5.21. Combined Feedback Linearization and MPC for Wind Turbine Power Tracking

Participants: Guillaume Sandou, Nicolo Gionfra [CentraleSupélec], Houria Siguerdidjane [CentraleSupélec], Damien Faille [EDF], Philippe Loevenbruck [CentraleSupélec].

The problem of controlling a variable-speedvariable-pitch wind turbine in non conventional operating points is addressed.We aim to provide a control architecture for a general active power tracking problem for the entire operating envelope. The presented control enables to cope with system non linearities while handling state and input constraints, and avoiding singular points. Simulations are carried out based on a 600 kW turbine parameters. Montecarlo simulation shows that the proposed controller presents a certain degree of robustness with respect to the system major uncertainties [36].

5.22. Hierarchical Control of a Wind Farm for Wake Interaction Minimization

Participants: Guillaume Sandou, Nicolo Gionfra [CentraleSupélec], Houria Siguerdidjane [CentraleSupélec], Damien Faille [EDF], Philippe Loevenbruck [CentraleSupélec].

The problem of controlling a wind farm for power optimization by minimizing the wake interaction among wind turbines is addressed. We aim to evaluate the real gain in farm power production when the dynamics of the controlled turbines are taken into account. The proposed local control enables the turbines to track the required power references in the whole operating envelope, and under the major uncertainties of the system. Simulations are carried out based on a wind farm of 600 kW turbines and they show the actual benefit of considering the wake effect in the optimization algorithm [54].

5.23. Control of a model of chemostat with delay

Participants: Frederic Mazenc, Michael Malisoff [LSU], Jerome Harmand [INRA].

We provided in [39] a new control design for models of chemostats, under constant substrate input concentrations, using piecewise constant delayed measurements of the substrate concentration. The growth functions can be uncertain and are not necessarily monotone. The dilution rate is the control. We used a new Lyapunov approach to derive conditions on the largest sampling interval and on the delay length to ensure asymptotic stabilization properties of a componentwise positive equilibrium point.

5.24. Mathematical Modelling of Acute Myeloid Leukemia

Participants: Catherine Bonnet, Jean Clairambault [MAMBA project-team], François Delhommeau [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Walid Djema, Emilia Fridman [Tel-Aviv University], Pierre Hirsch [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Frédéric Mazenc.

ALMA project focuses on analysis of healthy and unhealthy blood cell production. Dynamics of cell populations are modeled and mathematically analyzed in order to explain why some pathological disorders may occur. The challenging problem that we are facing is to steadily extend both modelling and analysis aspects to constantly better represent this complex physiological mechanism, which is not yet fully understood. This year, we have progressed on this line [35] and particular emphasis has been placed on a new generation of differential systems, coupled to algebraic equations, modeling abnormal proliferation as observed in acute myeloid leukemia [65]. We have developed, in [34], Lyapunov-like techniques in order to derive global or local exponential stability conditions for that class of differential-difference hematopoietic models. A new model describing the coexistence between ordinary and mutated hematopoietic stem cells was introduced and analyzed in [33]. Above all, this was about giving theoretical conditions to guarantee the survival of healthy cells while eradicating unhealthy ones. Interpretation of mathematical results leads us to provide possibly innovative therapies by combining drugs infusions. By continuing on the path of models coupling healthy and malignant cells, we proposed a framework to investigate the phenomena of tumour dormancy, which goes beyond leukemias, to cover all types of cancer. Finally, in a recent study, we highlighted the role played by growth factors -hormone-like molecules- on the regulation of various biological features involved in hematopoietic mechanisms; that we interpret in the framework of switching systems with distributed delays.

5.25. Ananlysis of Dengue Fever SIR Model with time-varying parameters

Participants: Stefanella Boatto [Univ Feder Rio de Janeiro], Catherine Bonnet, Frédéric Mazenc.

Dengue fever is an infectious viral disease occuring in humans that is prevalent in parts of Central and South America, Africa, India and South-east Asia and which causes 390 millions of infections worldwilde. We have considered here a SIR model of Dengue fever with a periodically time-dependent infection rate. Such a model has been considered by other authors before but we focused here on different aspects such as the existence of a periodic stable orbit and the importance of the phase of the infection rates.

GAMMA3 Project-Team

4. New Results

4.1. Remaillage adaptatif pour la mise en forme de tôles minces et de composites

Participants: Laurence Moreau [correspondant], Abel Cherouat, Houman Borouchaki.

Au cours des simulations numériques de mise en forme en 3D, les grandes déformations mises en jeu font que le maillage subit de fortes distorsions. Il est alors nécessaire de remailler continuellement la pièce afin de pouvoir capturer les détails géométriques des surface en contact, adapter la taille du maillage à la solution physique et surtout pouvoir effectuer la simulation jusqu'à la fin du procédé de mise en forme. Lorsque la pièce est comprise entre des outils rigides (cas de l'emboutissage), aux problèmes de remaillage s'ajoutent aussi des difficultés sur la gestion du contact entre les pièces. Une méthode couplant une stratégie de remaillage adaptatif et une technique de projection a été développée. Afin de pouvoir réaliser des simulations numériques de composites tissés, une procédure spécifique a été ajoutée au remailleur afin de pouvoir raffiner les éléments finis bi-composants (association d'éléments finis de barre et de membrane orientés matérialisant le comportement de fibres chaîne et trame).

Ce travail a donnée lieu à 1 article.

4.2. Le formage incrémental : étude expérimentale, numérique et remaillage adaptatif

Participants: Laurence Moreau [correspondant], Abel Cherouat, Houman Borouchaki.

Le formage incrémental est un procédé de mise en forme récent permettant de mettre en forme des tôles minces grâce au déplacement d'un outil hémisphérique dont la trajectoire est pilotée par une machine à commande numérique. Ce procédé peu couteux est une alternative intéressante à l'emboutissage traditionnel pour les entreprises réalisant des pièces de petite taille à usage unique ou en petite série comme les entreprises biomédicales (prothèses, implants personnalisés..). Cependant, il reste encore des développements importants sur le plan numérique et expérimental pour que ce procédé soit industrialisable : problèmes d'état de surface, de non-respect de la géométrie, risques de rupture. Nous avons étudié numériquement et expérimentalement ce procédé de formage incrémental : développement d'une méthode de remaillage adaptée à ce procédé, optimisation des paramètres du procédé, étude du formage incrémental à chaud, étude du formage incrémental robotisé.r

Ce travail a donnée lieu à 2 articles et 5 participations à des conférences internationales.

4.3. Reconstruction de surface 3D à partir d'images numériques 2D

Participants: Laurence Moreau [correspondant], Abel Cherouat, Houman Borouchaki.

Ces travaux portent sur la reconstruction 3D d'objets à partir de plusieurs photos prises via des caméras calibrées avec des points de vue différents couvrant la totalité de la surface de l'objet. La méthodologie générale consiste à apparier les pixels correspondants de deux photos et obtenir des positions 3D via une technique de triangulation.. L'idée originale réside dans une nouvelle méthodologie automatique d'appariement de pixels. Elle comprend trois étapes : un motif présentant un maillage triangulaire aléatoire est projeté sur l'objet 3D, le maillage est identifié sur chaque photo et la technique de triangulation est appliquée aux sommets de ce maillage. La méthodologie de reconstruction 3D a été appliquée à la modélisation géométrique du buste féminin afin d'envisager des simulations de comportements statique et dynamique de ce buste. Ces travaux ont conduit aussi à la conception et la réalisation d'une cabine d'acquisition permettant de prendre 24 prises de vue de manière simultanée depuis un ordinateur extérieur à la cabine.

Ce travail a donnée lieu à 1 article et 2 participations à des conférences internationales.

4.4. Modélisation numérique, remaillage adaptatif et optimisation pour la morphologie de nanofils

Participants: Laurence Moreau [correspondant], Thomas Grosges.

L'objectif était de développer une méthode permettant de détecter et d'analyser la présence de nanomatériaux dans l'eau. Une voie possible consiste à étudier les effets liés aux couplages lumière-matière, c'est-à-dire la réponse photo-thermique des nanomatériaux illuminés par une onde électromagnétique. La méthode proposée consiste à étudier la réponse thermique du nanofil immergé sous l'illumination et à la relier à la bulle produite. Le problème multi physique est modélisé par un système d'équations couplées : équation de Helmoltz et équation de la chaleur. La résolution numérique de ces équations est effectuée par une méthode des éléments finis et un processus d'optimisation incluant des boucles de remaillages adaptatifs afin de contrôler la précision de la solution et assurer la convergence. Une étude de la morphologie de la bulle a été réalisée en fonction de paramètres géométriques et physiques. Deux fonctions permettant de relier la taille de la bulle à la taille et la forme du nanomatériau ont été définies. La résolution du modèle inverse, associé à ces fonctions, permettant de remonter à la morphologie du nanomatériau via celle de la bulle. L'efficacité et la pertinence du modèle ont été montrées en confrontant les résultats numériques aux résultats expérimentaux.

Ce travail a donnée lieu à 3 articles et 2 participations à des conférences internationales

4.5. Les outils de remaillage dans la simulation multi-physiques pour la fiabilisation des systèmes complexes

Participants: Abel Cherouat [correspondant], Houman Borouchaki.

Le projet concerne la maîtrise des outils de simulation numérique multi-physique avec remaillage adaptatif 3D pour la prévention de la fiabilité des systèmes complexes. Les systèmes étudiés sont des structures comportant des composants et des architectures mécaniques. Ils sont fortement contraints car la partie électro-magnétique est très sensible aux vibrations, variations et dilatations thermiques, et agressions physico-chimiques qui existent habituellement dans les systèmes mécaniques. La fiabilisation représente des enjeux majeurs pour ces systèmes. L'objectif final est d'étudier la fiabilisation de ces systèmes par des approches hybrides qui combinent les outils de simulation éléments finis multi-physiques couplées avec adaptation en temps réel des maillages éléments finis en 3D. L'analyse de la fiabilité et la synthèse pouvant être appliquées en cas de défaillance pour maintenir l'exploitation des systèmes.

Ce travail a donnée lieu à 4 articles et 2 participations à des conférences internationales.

4.6. Les matériaux innovants : mousses métalliques - AMF, textiles techniques, composites et agro-composites : Modélisation mécanique, Simulation avec remaillage, Reconstitution 3D et Modélisation géométrique

Participants: Abel Cherouat [correspondant], Houman Borouchaki, Shijie Zhu, Antony Sheedev.

Le contexte de l'étude sur les mousses est la modélisation du comportement mécanique, la reconstitution 3D de la morphologie des mousses à partir d'images tomographiques ou de la CAO géométrique, de l'optimisation et de la simulation de la déformation de mousse (métalliques ou AMF).

Le contexte de l'étude sur les agro-composites est la maitrise des matières naturelles, l'allégement des structures et la valorisation de l'émergence des textiles biodégradables pour des applications industrielles. Les investigations concernent les aspects d'élaboration et mise en œuvre des textiles secs ou pré-imprégnés (tissés, UD cousu et mats), de caractérisation-modélisation comportementale multi-échelle et de mise au point d'outils d'aide à la décision et d'éco-conception des matériaux fonctionnels.

Le contexte de l'étude sur les composites est l'éco-réparation *in-situ* des structures industrielles intégrant l'hybridation de procédés émergents d'*additive manufacturing* et le frittage micro-onde avec l'utilisation de nouvelles résines ou nuances de matériaux, la numérisation 3D, l'impression ou collage par balayage et le contrôle non destructif.

Le contexte de l'étude sur les tissus biologiques est le développement de méthode d'obtention des paramètres mécaniques des tissus vivants et des informations pour l'amélioration des prothèses post-chirurgicale (un sein artificiel). Une approche médicale de la modélisation du sein et de sa déformabilité a pour objectif de prédire les déformations des tissus pendant les interventions en tenant compte des constituants (graisses, glandes, peau et ligaments), mais ne concerne pas le comportement du sein et son remodelage par le bonnet ou son comportement pendant le sport.

Ce travail a donnée lieu à 11 articles et 8 participations à des conférences internationales.

4.7. Reconstruction 3D à partir d'image vs Scanner 3D, Maillage adaptatif par vision embarquée sur drones autonomes

Participants: Abel Cherouat [correspondant], Houman Borouchaki.

Dans le cadre de ce projet, on se propose de concevoir un système de reconstruction adaptative et temps réel de scènes 3D en se basant uniquement sur le flux d'images captées par une caméra embarquée sur un drone autonome. Un nuage de points peut être ainsi obtenu en traitant d'une manière efficace et temps-réel le flux d'images issues de la caméra mobile. Le nuage de points en temps réel est utilisé pour reconstruire les surfaces des objets constituant la scène, et surtout de quantifier la qualité de la reconstruction en fonction de la géométrie de ces surfaces. Les applications concernées sont les automates industriels, l'imagerie médicale, le *Smart Tracking*, la surveillance et la sécurité, la rétro-conception et la réalité augmentée, ...

Ce travail a donnée lieu à 2 articles et 2 participations à des conférences internationales.

4.8. Les outils de remaillage dans la simulation et l'optimisation de la mise en forme des matériaux

Participants: Abel Cherouat [correspondant], Houman Borouchaki, Laurence Moreau.

L'objectif scientifique de ce projet est de développer des modèles théoriques, numériques et géométriques nécessaires à la mise au point de méthodologies de simulation numérique et d'optimisation de procédés de fabrication et de mise en forme de composants et de structures mécaniques en petites ou en grandes déformations. Une attention particulière est accordée à la génération de maillage, de remaillage et de maillage adaptatif isotrope et anisotrope plan (2D), surfacique (2,5D) et volumique (3D), ainsi que des méthodes d'optimisation de maillages (en particulier surfacique) ainsi que les couplages multi-physiques entre les différents phénomènes.

Ce travail a donnée lieu à 13 articles et 4 participations à des conférences internationales.

4.9. Applications du maillage et développements de méthodes avancées pour la cryptographie

Participants: Thomas Grosges [correspondant], Dominique Barchiesi, Michael François.

L'utilisation des nombres (pseudo)-aléatoires a pris une dimension importante ces dernières décennies. De nombreuses applications dans le domaine des télécommunications, de la cryptographie, des simulations numériques ou encore des jeux de hasard, ont contribué au développement et à l'usage de ces nombres. Les méthodes utilisées pour la génération de tels nombres (pseudo)-aléatoires proviennent de deux types de processus : physique et algorithmique. Ce projet de recherche a donc pour objectif principal le développement de nouveaux procédés de génération de clés de chiffrement, dits "exotiques", basés sur des processus physiques, multi-échelles, multi-domaines assurant un niveau élevé de sécurité. Deux classes de générateurs basés sur des principes de mesures physiques et des processus mathématiques ont été développé.

La première classe de générateurs exploite la réponse d'un système physique servant de source pour la génération des séquences aléatoires. Cette classe utilise aussi bien des résultats de simulations que des résultats de mesures interférométriques pour produire des séquences de nombres aléatoires. L'application du maillage adaptatif sert au contrôle de l'erreur sur la solution des champs physiques (simulés ou mesurés). À partir de ces cartes physiques, un maillage avec estimateur d'erreur sur l'entropie du système est appliqué. Celui-ci permet de redistribuer les positions spatiales des noeuds. L'étude (locale) de la réduction d'entropie des clés tout au long de la chaîne de création et l'étude (globale) de l'entropie de l'espace des clés générées sont réalisées à partir de tests statistiques.

La seconde classe de générateurs porte sur le développement de méthodes avancées et est basée sur l'exploitation de fonctions chaotiques en utilisant les sorties de ces fonctions comme indice de permutation sur un vecteur initial. Ce projet s'intéresse également aux systèmes de chiffrement pour la protection des données et deux algorithmes de chiffrement d'images utilisant des fonctions chaotiques sont développés et analysés. Ces algorithmes utilisent un processus de permutation-substitution sur les bits de l'image originale. Une analyse statistique approfondie confirme la pertinence des cryptosystèmes développés. Les résultats de cette recherche se sont vu récompensés par un premier prix décerné par EURASIP (European Association in Signal Processing) en 2016 ("Best paper award of the EURASIP).

4.10. Méthodes avancées pour la nanomorphologie des nanotubes/fils en suspension liquide"

Participants: Thomas Grosges [correspondant], Dominique Barchiesi, Abel Cherouat, Houman Borouchaki, Laurence Giraud-Moreau, Anis Chaari.
Validité du projet: 2011-2015.
Production scientifique: 1 thèse soutenue en 2016 (A. Chaari), 3 articles publiés, 1 conférence nternationale (PIERS 2014), 2 conférences nationales (CSMA 2013 et CSMA 2015).
Ce projet de recherche (NANOMORPH) a pour objet principal le développement et la mise au point d'une instrumentation optique pour déterminer la distribution en tailles et le coefficient de forme de nanofils (NF) ou de nanotubes (NT) en suspension dans un écoulement. Au cours de ce projet, deux types de techniques optiques complémentaires sont développées. La première, basée sur la diffusion statique de la lumière, nécessite d'étudier au préalable la physico-chimie de la dispersion, la stabilisation et l'orientation des nanofils dans les milieux d'étude. La seconde méthode, basée sur une méthode opto-photothermique pulsée, nécessite en sus, la modélisation de l'interaction laser/nanofils, ainsi que l'étude des phénomènes multiphysiques induits par ce processus. L'implication de l'équipe-projet GAMMA3 concerne principalement la simulation multiphysique de l'interaction laser-nanofils et l'évolution temporelle des bulles et leurs formations. L'une des principales difficultés de ces problématiques est que la géométrie du domaine est variable (à la fois au sens géométrique et topologique). Ces simulations ne peuvent donc être réalisées que dans un schéma adaptatif de calcul nécessitant le remaillage tridimensionnel mobile, déformable avec topologie variable du domaine (formation et évolution des bulles au cours du temps et de l'espace).

4.11. Méthodes de résolutions avancées et modélisation électromagnetisme-thermique-mécanique à l'échelle mesoscopique

Participants: Dominique Barchiesi [correspondant], Abel Cherouat, Thomas Grosges, Houman Borouchaki, Laurence Giraud-Moreau, Sameh Kessentini, Anis Chaari, Fadhil Mezghani
Validité du projet: 2009-2016 (thèse de Fadhil Mezghani initiée en 2012 coencadrée par D. Barchiesi et A. Cherouat).
Production scientifique: 2 thèses soutenues (S. Kessentini, 22/10/2012 et F. Mezghani), 15 articles publiés, 6 conférences.

Le contrôle et l'adaptation du maillage lors de la résolution de problèmes couplés et/ou non linéaires reste un problème ouvert et fortement dépendant du type de couplage physique entre les EDP à résoudre. Notre objectif est de développer des modèles stables afin de calculer les dilatations induites par l'absorption d'énergie électromagnétique, par des structures matérielles inférieures au micron. Les structures étudiées sont en particulier des nanoparticules métalliques en condition de résonance plasmon. Dans ce cas, un maximum d'énergie absorbée est attendu, accompagné d'un maximum d'élévation de température et de dilatation. Il faut en particulier développer des modèles permettant de simuler le comportement multiphysique de particules de formes quelconques, pour une gamme de fréquences du laser d'éclairage assez étendue afin d'obtenir une étude spectroscopique de la température et de la dilatation. L'objectif intermédiaire est de pouvoir quantifier la dilatation en fonction de la puissance laser incidente. Le calcul doit donc être dimensionné et permettre finalement des applications dans les domaines des capteurs et de l'ingénierie biomédicale. En effet, ces nanoparticules métalliques sont utilisées à la fois pour le traitement des cancers superficiels par nécrose de tumeur sous éclairage adéquat, dans la fenêtres de transparence cellulaire. Déposées sur un substrat de verre, ces nanoparticules permettent de construire des capteurs utilisant la résonance plasmon pour être plus sensibles (voir projet européen *Nanoantenna* et l'activité génération de nombres aléatoires). Cependant, dans les deux cas, il est nécessaire, en environnement complexe de déterminer la température locale, voire la dilatation de ces nanoparticules, pouvant conduire à un désaccord du capteur, la résonance plasmon étant très sensible aux paramètres géométriques et matériels des nanostructures. En ce sens, l'étude permet d'aller plus loin que la "simple" interaction électromagnétique avec la matière du projet européen *Nanoantenna*.

Le travail a constitué en la poursuite de l'étude des spécificités de ce type de problème multiphysique pour des structures de forme simple et la mise en place de fonctions test, de référence, pour les développements de maillage adaptatifs pour les modèles multiphysiques éléments finis. Nous espérons pouvoir proposer un projet ANR couplant les points de vue microscopiques et macroscopiques dans les prochaines années.

4.12. Problèmes de magnétostatique sur maillage de grande taille et multi-échelle

Participants: Dominique Barchiesi [correspondant], Thomas Grosges, Houman Borouchaki, Brahim Yahiaoui
Validité du projet: 2013-2015. Post-doc Brahin Yahiaoui.
Le projet Flyprod concerne l'étude du stockage d'électricité par volant d'inertie lévité et financé par l'ADEME. Une technologie brevetée innovante et stratégique permettant à des acteurs majeurs de la distribution électrique de stocker de l'énergie pour des périodes de fortes consommations. D'un point de vue écologique, un volant d'inertie n'émet ni gaz à effet de serre, ni produits chimiques nocifs pour l'environnement. Les partenaires pour ce projet sont LEVISYS, Université de Technologie de Troyes, SCLE SFE (COFELY INEO, Groupe GDFSUEZ), CIRTEM, Conseil Général de l'Aube. Les dispositifs mis en oeuvre nécessitent des études approfondies pour rendre les volants d'inertie économiquement viables. La recherche a consisté à développer un programme informatique permettant une simulation assistée par ordinateur. Il permet plus précisément de calculer les champs magnétiques et de concevoir les pièces du volant d'inertie afin de garantir une perte minimale d'énergie. Le champ magnétique doit être calculé en un temps raisonnable sur des distances spatiales réduites. L'approche utilisée pour répondre à ces objectifs est appliquée sur un maillage fourni par le logiciel Optiform (un remailleur adaptatif volumique développé par l'équipe GAMMA3). Les résultats obtenus ont permis d'optimiser la structure du volant d'inertie et d'atteindre une efficacité de stockage de 97%, permettant de valider la pertinence du volant et de confirmer sa fabrication.

4.13. Element metric, element quality and interpolation error metric

Participants: Paul Louis George [correspondant], Houman Borouchaki.

The metric of a simplex of $\mathbb{R}^d$ is a metric tensor (symmetric positive definite matrix) in which the element is unity (regular with unit edge lengths). This notion is related to the problem of interpolation error of a given field over a mesh. Let K be a simplex and let us denote by v_{ij} the vector joining vertex i and vertex j of K. The metric of K can be written as:

$$\mathcal{M} = \frac{d+1}{2}(\sum_{i<j} v_{ij}\,{}^t v_{ij})^{-1},$$

where $v_{ij}\,{}^t v_{ij}$ is a $d \times d$ rank 1 matrix related to edge ij.

The metric of a simplex also characterizes the element shape. In particular, if it is the identity, the element is unity. Hence, to define the shape quality of an element, one can determine the gap of the element metric $\mathcal{M}$ and the identity using different measures based on the eigenvalues $\lambda_i = \frac{1}{h_I^2}$ of $\mathcal{M}$ or those of $\mathcal{M}^{-1}$, e.g. h_i^2. Notice that metric $\mathcal{M}^{-1}$ is directly related to the geometry of the element (edge length, facet area, element volume). The first algebraic shape quality measure ranging from 0 to 1 is defined as the ratio of the geometric average of the eigenvalues of $\mathcal{M}^{-1}$ and their arithmetic average:

$$q(K) = \frac{\left(\prod_i h_i^2\right)^{\frac{1}{d}}}{\frac{1}{d}\sum_{i=1}^{d} h_i^2} = d\,\frac{\left(det(\mathcal{M}^{-1})\right)^{\frac{1}{d}}}{tr(\mathcal{M}^{-1})}.$$

As the geometric average is smaller than the arithmetic average, this measure is well defined. In addition, it is the algebraic reading of the well-known quality measure defined by:

$$q^{\frac{d}{2}}(K) = (d!)d^{\frac{d}{2}}(d+1)^{\frac{d-1}{2}}\frac{|K|}{\left(\sum_{i<j} l_{ij}^2\right)^{\frac{d}{2}}},$$

where the volume and the square of the edge lengths are involved. The algebraic meaning justifies the above geometric measure. The second algebraic shape quality measure is defined as the ratio of the harmonic average of the eigenvalues of $\mathcal{M}^{-1}$ and their arithmetic average (ranging also from 0 to 1):

$$q(K) = \frac{\left\{\frac{1}{d}\sum_{i=1}^{d}\frac{1}{h_i^2}\right\}^{-1}}{\frac{1}{d}\sum_{i=1}^{d} h_i^2} = \frac{d^2}{tr(\mathcal{M})tr(\mathcal{M}^{-1})}.$$

As above, this measure is well defined, the harmonic average being smaller the arithmetic one. From this measure, one can derive another well-known measure involving the roundness and the size of an element (measure which is widely used for convergene issues in finite element methods).

Note that these measures use the invariants of $\mathcal{M}^{-1}$ or $\mathcal{M}$ and thus can be evaluated from the coefficients of the characteristical polynomial of those matrices (avoiding the effective calculation of their eigenvalues). Another advantage of the above algebraic shape measures is their easy extensions in an arbitrary Euclidean space. Indeed, if $\mathcal{E}$ is the metric of such a space, the algebraic shape measures read:

$$q_{\mathcal{E}}(K) = d\,\frac{\left(det(\mathcal{M}^{-1}\mathcal{E})\right)^{\frac{1}{d}}}{tr(\mathcal{M}^{-1}\mathcal{E})} \quad , \quad q_{\mathcal{E}}(K) = \frac{d^2}{tr(\mathcal{E}^{-1}\mathcal{M})tr(\mathcal{M}^{-1}\mathcal{E})}.$$

This work has been published in a journal, [8].

Following this notion of a element metric, a natural work was done regarding how to define the element metric so as to achieve a given accuracy for the interpolation error of a function using a finite element approximation by means of simplices of arbitrary degree.

This is a new approach for the majoration of the interpolation error of a polynomial function of arbitrary degree n interpolated by a polynomial function of degree $n-1$. From that results a metric, the so-called interpolation metric, which allows for a control of the error. The method is based on the geometric and algebraic properties of the metric of a given element, metric in which the element is regular and unit. The interpolation metric plays an important role in advanced computations based on mesh adaptation. The method relies in a Bezier reading of the functions combined with Taylor expansions. In this way, the error in a given element is fully controled at the time the edges of the element are controled.

It is shown that the error in bounded as

$$|e(X)| \leq C \sum_{i<j} f^{(n)}(.)(v_{ij}, v_{ij}, ..., v_{ij}),$$

where C is a constant depending on d and n, v_{ij} is the edge from the vertices of K of index i and j, $f^{(n)}(.)$ is the derivative of order n of f applied to a n-uple uniquely composed of v_{ij}. If we consider the case $d=2$ and $u=(x,y)$ is a vector in $\mathbb{R}^2$, we have

$$f^{(n)}(.)(u,u,...,u) = \sum_{i=0}^{n-2} x^{n-2-i} y^i \ {}^t u \left(C_i^{n-2} \mathcal{H}_{(n-2,n-2-i,i)} \right) u \ ,$$

where the quadratic forms $\mathcal{H}_{(n-2,n-2-i,i)}$ are defined by the matrices of order 2 (with constant entries):

$$\mathcal{H}_{(n-2,n-2-i,i)} = \begin{pmatrix} \frac{\partial^{(n)} f}{\partial x_1^{n-i} \partial x_2^i} & \frac{\partial^{(n)} f}{\partial x_1^{n-1-i} \partial x_2^{i+1}} \\ \\ \frac{\partial^{(n)} f}{\partial x_1^{n-1-i} \partial x_2^{i+1}} & \frac{\partial^{(n)} f}{\partial x_1^{n-2-i} \partial x_2^{i+2}} \end{pmatrix},$$

those matrices being the hessians of the derivatives of f of order $n-2$.

This work resulted in a paper submitted in a journal and currently under revision.

4.14. Realistic modeling of fractured geologic media

Participants: Patrick Laug [correspondant], Géraldine Pichot.

This study, in collaboration with project-team Serena, aims to model, in a realistic and efficient manner, natural fractured media. These media are characterized by their diversity of structures and organizations. Numerous studies in the past decades have evidenced the existence of characteristic structures at multiple scales. At fracture scale, the aperture distribution is widely correlated and heterogeneous. At network scale, the topology is complex resulting from mutual mechanical interactions as well as from major stresses. Geometric modeling of fractured networks combines in a non-standard way a large number of 2D fractures interconnected in the 3D space. Intricate local configurations of fracture intersections require original methods of geometric modeling and mesh generation. We have developed in 2016 a software package that automatically builds geometric models and surface meshes of random fracture networks. The results are highly promising and we now want to continue this research to further improve the element quality in complex configurations, take into account multiple size scales in large fracture networks (up to thousands of fractures), and compare several modeling strategies (mixed hybrid finite elements, projected grids, mortar elements) [13].

4.15. Parallel meshing of surfaces defined by collections of connected regions

Participant: Patrick Laug [correspondant].

In CAD (computer aided design) environments, a surface is commonly modeled as a collection of connected regions represented by parametric mappings. For meshing such a composite surface, a parallelized indirect approach with dynamic load balancing can be used on a shared memory system. However, this methodology can be inefficient in practice because most existing CAD systems use memory caches that are only appropriate to a sequential process. As part of the sabbatical year of P. Laug at Polytechnique Montréal in 2014/2015, two solutions have been proposed, referred to as the Pirate approach and the Discrete approach. In the first approach, the Pirate library can be efficiently called in parallel since no caching is used for the storage or evaluation of geometric primitives. In the second approach, the CAD environment is replaced by internal procedures interpolating a discrete geometric support. In 2016, the dynamic load balancing has been analyzed and improved. Significant modifications to the Pirate library have been made, and new numerical tests on three different computers (4, 8 and 64 cores) have been carried out, now showing an almost linear scaling of the method in all cases [10].

4.16. Discrete CAD model for visualization and meshing

Participants: Patrick Laug [correspondant], Houman Borouchaki.

During the design of an object using a CAD (computer aided design) platform, the user can visualize the ongoing model at every moment. Visualization is based on a discrete representation of the model that coexists with the exact analytical representation of the object. Most CAD systems have this discrete representation available, and each of them applies its own construction methodology. We have developed in 2016 a method to build a discrete model for CAD surfaces (the model is quadtree-based and subdivided into quadrilaterals and triangles). The method presents two major particularities: most elements are aligned with iso-parametric curves and the accuracy of the surface approximation is controlled. In addition, we have proposed a new technique of surface mesh generation that is based on this discrete model. This approach has been implemented as a part of a surface mesher called ALIEN, and several examples have demonstrate the robustness and computational efficiency of the program, as well as the quality of the geometric support [14], [15].

4.17. Visualization and modification of high-order curved meshes

Participants: Alexis Loyer, Dave Marcum, Adrien Loseille [correspondant].

During the partnership between Inria and Distene, a new visualization software has been designed. It address the typical operations that are required to quickly assess the newly algorithm developed in the team. In particular, interactive modifications of high-order curved mesh and hybrid meshes has been addressed. The software VIZIR is freely available at https://www.rocq.inria.fr/gamma/gamma/vizir/.

4.18. Adaptation de maillages pour des écoulements visqueux en turbomachine

Participants: Frédéric Alauzet, Loïc Frazza, Adrien Loseille [correspondant].

4.18.1. Calcul.

Les prémices d'une adaptation pour les écoulements Navier-Stokes turbulents ont été testés sur des calculs de turbomachine. Pour ce faire nous avons tout d'abord traité les particularités liées aux calculs en turbomachine: - Les aubes présentent en général une périodicité par rotation et on ne simule donc qu'une période afin d'alléger les calculs. Il faut donc traiter cette périodicité de façon appropriée dans le code CFD et l'adaptation de maillage. - Afin de prendre en compte la rotation des pales sans employer de maillages mobiles et simulations instationnaires on peut se placer dans le référentiel tournant de l'aube en corrigeant les équations. - Les écoulements en turbomachine sont des écoulements clos, les conditions limites d'entrée et de sortie ont donc une influence très forte et peuvent de plus se trouver très près de la turbine afin de simuler la présence d'autres étages en amont ou aval. Des conditions limites bien précises ont donc été développées afin de traiter correctement ces effets.

4.18.2. Adaptation.

Pour l'adaptation de maillages deux particularités doivent être traitées ici, la périodicité du maillage et la couche limite turbulente.

En 2D, la couche limite turbulente est automatiquement adaptée avec la méthode metric orthogonal et la périodicité du maillage est garantie par un traitement spécial des frontières. Les estimateurs d'erreurs Navier-Stokes et RANS n'étant pas encore au point nous avons utilisé la Hessienne du Mach de l'écoulement comme senseur ce qui donne déjà des résultats satisfaisants.

En 3D la méthode metric orthogonal est beaucoup plus complexe à mettre en oeuvre et n'est pas encore au point. La couche limite a donc été exclue de l'adaptation, le maillage est adapté uniquement dans le volume en utilisant la Hessienne du Mach de l'écoulement comme senseur. La périodicité n'étant pas traitée non plus, les frontières périodiques restent inchangées ce qui garantie leur périodicité.

4.18.3. Norm-Oriented.

Dans le cadre de la théorie Norm-Oriented, afin de contrôler l'erreur implicite des schémas numériques, un correcteur a été développé et testé. Etant donné un maillage et la solution numérique obtenue avec, le résidu de cette solution projeté sur un maillage deux fois plus fin est accumulé sur le maillage initial. Ce défaut de résidu est utilisé comme terme source dans une seconde simulation plus courte. La nouvelle solution toujours sur le meme maillage est plus proche de la solution exacte et donne une bonne estimation de l'erreur.

4.19. Metric-orthogonal and metric-aligned mesh adaptation

Participants: Frédéric Alauzet, Loïc Frazza, Adrien Loseille, Dave Marcum [correspondant].

A new algorithm to derive adaptive meshes has been introduced through new cavity-based algorithms. It allows to generate anisotropic surface and volume mesh that are aligned along the eigenvector directions. This allows us to improve the quality of the meshes and to deal naturally with boundary layer mesh generation.

Die orthogonale metric Methode erzeugt 2D-Elemente mit einem Rand, der mit der Hauptrichtung der Metrik ausgerichtet ist und einem zweiten Rand, der rechtwinklig zur ersten ist. Das erzeugende Gitter is so örtlich strukturiert wo es Anisotropie gibt. Dieses Methode wurde erfolgreich zur automatischen strukturierten Gitter Erzeugung in der turbulenten Grenzschichten für Turbomaschinen Simulationen angewendet.

4.20. Parallel mesh adaptation

Participants: Frédéric Alauzet, Adrien Loseille [correspondant].

We devise a strategy in order to generate large-size adapted anisotropic meshes $O(10^8 - 10^9)$ as required in many fields of application in scientific computing. We target moderate scale parallel computational resources as typically found in R&D units where the number of cores ranges in $O(10^2 - 10^3)$. Both distributed and shared memory architectures are handled. Our strategy is based on hierarchical domain splitting algorithm to remesh the partitions in parallel. Both the volume and the surface mesh are adapted simultaneously and the efficiency of the method is independent of the complexity of the geometry. The originality of the method relies on (i) a metric-based static load-balancing, (ii) dedicated hierarchical mesh partitioning techniques to (re)split the (complex) interfaces meshes, (iii) anisotropic Delaunay cavity to define the interface meshes, (iv) a fast, robust and generic sequential cavity-based mesh modification kernel, and (v) out-of-core storing of completing parts to reduce the memory footprint. We are able to generate (uniform, isotropic and anisotropic) meshes with more than 1 billion tetrahedra in less than 20 minutes on 120 cores [11].

4.21. Unsteady adjoint computation on dynamic meshes

Participants: Eléonore Gauci, Frédéric Alauzet [correspondant].

Adjoint formulations for unsteady problems are less common due to the extra complexity inherent in the numerical solution and storage but these methods are a great option in engineering because it takes more into account the cost function we want to minimize. Moreover the engineering applications involve moving bodies and this motion must be taken into account by the governing flow equations. We develop a model of unsteady adjoint solver on moving mesh problems. The derivation of the adjoint formulation based on the ALE form of the equations requires consideration of the dynamic meshes. Our model takes into account the DGCL.

4.22. Line solver for efficient stiff parse system resolution

Participants: Loïc Frazza, Frédéric Alauzet [correspondant].

Afin d'accélérer la résolution des problèmes raides, un line-solver á été développé. Cette méthode extrait tout d'abord des lignes dans le maillage du problème selon des critères géométriques ou physiques. Le problème peut alors être résolu exactement le long des ces lignes à moindre coût. Cette méthode est particulièrement bien adaptée aux cas où l'information se propage selon une direction privilégiée tels que les chocs, les couches limites ou les sillages. Ces cas sont généralement associés à des maillages très étirés ce qui conduit à des problèmes raides mais quasi-unidimensionnels. Ils peuvent donc être résolu efficacement par un line-solver, réduisant ainsi les temps de calculs tout en gagnant en robustesse.

4.23. Error estimate for high-order solution field

Participants: Olivier Coulaud, Adrien Loseille [correspondant].

Afin de produire des solveurs d'ordre élevé, et ainsi répondre aux exigences inhérentes à la résolution de problèmes physiques complexes, nous développons une méthode d'adaptation de maillage d'ordre élevé. Celle-ci est basée sur le contrôle par une métrique de l'erreur d'interpolation induite par le maillage du domaine. Plus précisément, pour une solution donnée, l'erreur d'interpolation d'ordre k est paramétrée par la forme différentielle $(k+1)^{\text{ième}}$ de cette solution, et le problème se réduit à trouver la plus grande ellipse incluse dans une ligne de niveau de cette différentielle. La méthode que nous avons mise au point théoriquement et numériquement est appelée "log-simplexe", et permet de produire des maillages adaptés d'ordre élevé dans un temps raisonnable, et ce en dimension 2 et 3. À l'occasion de l'International Meshing Roundtable 2016, ce travail a été présenté et publié. D'autres applications de cette méthode sont en cours d'exploitation, comme par exemple la génération de maillages adaptés courbes de surface, ou le couplage avec un solveur d'ordre élevé.

4.24. Méthode d'immersion de frontières pour la mécanique des fluides

Participants: Frédéric Alauzet [correspondant], Rémi Feuillet, Adrien Loseille.

Dans les méthodes de résolution classiques des problèmes d'interaction fluide-structure, il est usuel de représenter l'objet de manière exacte dans le maillage, c'est-à-dire avec des éléments conformes à l'objet : le maillage possède des triangles dont une arête correspond avec le bord de la géométrie immergée. Cette méthode quoique plus précise est très coûteuse en préprocessing. C'est dans ce cadre qu'est introduite la notion d'immersion de frontière (embedded geometry en anglais). Cette méthode consiste à représenter la géométrie de manière fictive. Le maillage de calcul n'est de fait plus nécessairement conforme à la géométrie de l'objet. Il s'agit donc de s'intéresser aux modifications nécessaires sur les méthodes classiques pour faire un calcul dans le cadre de l'immersion de frontières. Cela concerne les conditions aux limites et l'avancée en temps. On s'intéresse également à l'adaptation de maillage pour le cas de l'immersion. La finalité de tout ce travail est d'effectuer des calculs de coefficients aérodynamiques (portance, traînée) et de trouver des résultats du même ordre de précision que ceux en géométrie inscrite.

4.25. Optimisation de formes et CAO

Participants: Frédéric Alauzet [correspondant], Jean de Becdelièvre, Adrien Loseille.

Pour ce stage de 3 mois, l'objectif était de réaliser entièrement une optimisation aérodynamique, de la génération des modèles 3D aux calculs de la forme optimisée. Le modèle choisi était l'aile du C.R.M. (Common Research Model) de la NASA qui a été extensivement testé en soufflerie. Durand la première phase du projet, l'outils EGADS (Engineering Geometry Aircraft Design System) développé par le Aerospace Computational Design Lab (M.I.T) a été utilisé pour générer des modèles 3D paramétriques. À cette occasion, un outil facilement réutilisable de génération de modèle d'aile a été développé, ainsi que des outils de modification des modèles C.A.D. sous EGADS. Les maillages surfaciques de ces modèles ont été créés par EGADS directement et modifiés immédiatement par AMG pour les adapter au calcul. Les maillages volumiques ont, eux, été générés par GHS3D. Des calculs non visqueux sur des maillages adaptés ont alors permis d'obtenir des résultats, et de répéter l'opération jusqu'à obtenir un minimum. L'originalité de cette optimisation est que chaque calcul, à chaque itération de l'optimiseur, utilise un maillage adapté à l'aide des solutions des calculs précédents ; ce qui permet de réduire les coûts de calcul et d'augmenter la précision.

4.26. Boundary layer mesh generation

Participants: Frédéric Alauzet [correspondant], Adrien Loseille, Dave Marcum.

A closed advancing-layer method for generating high-aspect-ratio elements in the boundary layer (BL) region has been developed. This approach provides an answer to the mesh generation robustness issue as it starts from an existing valid mesh and always guarantuees a valid mesh in output. And, it handles very efficiently and naturally BL front collisions and it produces a natural smooth anisotropic blending between colliding layers. In addition, it provides a robust strategy to couple unstructured anisotropic mesh adaptation and high-aspect-ratio element pseudo-structured BL meshes. To this end, the mesh deformation is performed using the metric field associated with the given anisotropic meshes to maintain the adaptivity while inflating the BL. This approach utilizes a recently developed connectivity optimization based moving mesh strategy for deforming the volume mesh as the BL is inflated. In regards to the BL mesh generation, it features state-of-art capabilities, including, optimal normal evaluation, normal smoothing, blended BL termination, mixed-elements BL, varying growth rate, and BL imprinting on curved surfaces. Results for typical aerospace configurations are presented to assess the proposed strategy on both simple and complex geometries.

GECO Project-Team

7. New Results

7.1. New results: geometric control

Let us list some new results in sub-Riemannian geometry and hypoelliptic diffusion obtained by GECO's members.

- In [2] we compare different notions of curvature on contact sub-Riemannian manifolds. In particular we introduce canonical curvatures as the coefficients of the sub-Riemannian Jacobi equation. The main result is that all these coefficients are encoded in the asymptotic expansion of the horizontal derivatives of the sub-Riemannian distance. We explicitly compute their expressions in terms of the standard tensors of contact geometry. As an application of these results, we obtain a sub-Riemannian version of the Bonnet-Myers theorem that applies to any contact manifold.
- In [3] we provide the small-time heat kernel asymptotics at the cut locus in three relevant cases: generic low-dimensional Riemannian manifolds, generic 3D contact sub-Riemannian manifolds (close to the starting point) and generic 4D quasi-contact sub-Riemannian manifolds (close to a generic starting point). As a byproduct, we show that, for generic low-dimensional Riemannian manifolds, the only singularities of the exponential map, as a Lagragian map, that can arise along a minimizing geodesic are A_3 and A_5 (in Arnol'd's classification). We show that in the non-generic case, a cornucopia of asymptotics can occur, even for Riemannian surfaces.
- In [5] we study the evolution of the heat and of a free quantum particle (described by the Schrödinger equation) on two-dimensional manifolds endowed with the degenerate Riemannian metric $ds^2 = dx^2 + |x|^{-2\alpha} d\theta^2$, where $x \in \mathbb{R}$, $\theta \in S^1$ and the parameter $\alpha \in \mathbb{R}$. For $\alpha \leq -1$ this metric describes cone-like manifolds (for $\alpha = -1$ it is a flat cone). For $\alpha = 0$ it is a cylinder. For $\alpha \geq 1$ it is a Grushin-like metric. We show that the Laplace-Beltrami operator Δ is essentially self-adjoint if and only if $\alpha \notin (-3, 1)$. In this case the only self-adjoint extension is the Friedrichs extension Δ_F, that does not allow communication through the singular set $\{x = 0\}$ both for the heat and for a quantum particle. For $\alpha \in (-3, -1]$ we show that for the Schrödinger equation only the average on θ of the wave function can cross the singular set, while the solutions of the only Markovian extension of the heat equation (which indeed is Δ_F) cannot. For $\alpha \in (-1, 1)$ we prove that there exists a canonical self-adjoint extension Δ_N, called bridging extension, which is Markovian and allows the complete communication through the singularity (both of the heat and of a quantum particle). Also, we study the stochastic completeness (i.e., conservation of the L^1 norm for the heat equation) of the Markovian extensions Δ_F and Δ_B, proving that Δ_F is stochastically complete at the singularity if and only if $\alpha \leq -1$, while Δ_B is always stochastically complete at the singularity.
- In [6] we study spectral properties of the Laplace–Beltrami operator on two relevant almost-Riemannian manifolds, namely the Grushin structures on the cylinder and on the sphere. As for general almost-Riemannian structures (under certain technical hypothesis), the singular set acts as a barrier for the evolution of the heat and of a quantum particle, although geodesics can cross it. This is a consequence of the self-adjointness of the Laplace–Beltrami operator on each connected component of the manifolds without the singular set. We get explicit descriptions of the spectrum, of the eigenfunctions and their properties. In particular in both cases we get a Weyl law with dominant term $E \log E$. We then study the effect of an Aharonov-Bohm non-apophantic magnetic potential that has a drastic effect on the spectral properties. Other generalized Riemannian structures including conic and anti-conic type manifolds are also studied. In this case, the Aharonov-Bohm magnetic potential may affect the self-adjointness of the Laplace-Beltrami operator.

- Generic singularities of line fields have been studied for lines of principal curvature of embedded surfaces. In [7] we propose an approach to classify generic singularities of general line fields on 2D manifolds. The idea is to identify line fields as bisectors of pairs of vector fields on the manifold, with respect to a given conformal structure. The singularities correspond to the zeros of the vector fields and the genericity is considered with respect to a natural topology in the space of pairs of vector fields. Line fields at generic singularities turn out to be topologically equivalent to the Lemon, Star and Monstar singularities that one finds at umbilical points.
- In [10] we prove that any corank 1 Carnot group of dimension $k+1$ equipped with a left-invariant measure satisfies the measure contraction property $\mathrm{MCP}(K,N)$ if and only if $K \leq 0$ and $N \geq k+3$. This generalizes the well known result by Juillet for the Heisenberg group H^{k+1} to a larger class of structures, which admit non-trivial abnormal minimizing curves. The number $k+3$ coincides with the geodesic dimension of the Carnot group, which we define here for a general metric space. We discuss some of its properties, and its relation with the curvature exponent (the least N such that the $\mathrm{MCP}(0,N)$ is satisfied). We prove that, on a metric measure space, the curvature exponent is always larger than the geodesic dimension which, in turn, is larger than the Hausdorff one. When applied to Carnot groups, our results improve a previous lower bound due to Rifford. As a byproduct, we prove that a Carnot group is ideal if and only if it is fat.
- In [14] we relate some basic constructions of stochastic analysis to differential geometry, via random walk approximations. We consider walks on both Riemannian and sub-Riemannian manifolds in which the steps consist of travel along either geodesics or integral curves associated to orthonormal frames, and we give particular attention to walks where the choice of step is influenced by a volume on the manifold. A primary motivation is to explore how one can pass, in the parabolic scaling limit, from geodesics, orthonormal frames, and/or volumes to diffusions, and hence their infinitesimal generators, on sub-Riemannian manifolds, which is interesting in light of the fact that there is no completely canonical notion of sub-Laplacian on a general sub-Riemannian manifold. However, even in the Riemannian case, this random walk approach illuminates the geometric significance of Ito and Stratonovich stochastic differential equations as well as the role played by the volume.
- By adapting a technique of Molchanov, we obtain in [15] the heat kernel asymptotics at the sub-Riemannian cut locus, when the cut points are reached by a r-dimensional parametric family of optimal geodesics. We apply these results to the bi-Heisenberg group, that is, a nilpotent left-invariant sub-Riemannian structure on $\mathbb{R}^5$ depending on two real parameters α_1 and α_2. We develop some results about its geodesics and heat kernel associated to its sub-Laplacian and we point out some interesting geometric and analytic features appearing when one compares the isotropic ($\alpha_1 = \alpha_2$) and the non-isotropic cases ($\alpha_1 \neq \alpha_2$). In particular, we give the exact structure of the cut locus, and we get the complete small-time asymptotics for its heat kernel.
- The Whitney extension theorem is a classical result in analysis giving a necessary and sufficient condition for a function defined on a closed set to be extendable to the whole space with a given class of regularity. It has been adapted to several settings, among which the one of Carnot groups. However, the target space has generally been assumed to be equal to $\mathbb{R}^d$ for some $d \geq 1$. We focus in [17] on the extendability problem for general ordered pairs (G_1, G_2) (with G_2 non-Abelian). We analyze in particular the case $G_1 = \mathbb{R}$ and characterize the groups G_2 for which the Whitney extension property holds, in terms of a newly introduced notion that we call pliability. Pliability happens to be related to rigidity as defined by Bryant an Hsu. We exploit this relation in order to provide examples of non-pliable Carnot groups, that is, Carnot groups so that the Whitney extension property does not hold. We use geometric control theory results on the accessibility of control affine systems in order to test the pliability of a Carnot group.
- In [19] we study the cut locus of the free, step two Carnot groups G^k with k generators, equipped with their left-invariant Carnot–Carathéodory metric. In particular, we disprove the conjectures on the shape of the cut loci proposed in the literature, by exhibiting sets of cut points $C \subset G^k$ which, for $k \geq 4$, are strictly larger than conjectured ones. Furthermore, we study the relation of the cut locus with the so-called abnormal set. For each $k \geq 4$, we show that, contrarily to the case $k = 2, 3$,

the cut locus always intersects the abnormal set, and there are plenty of abnormal geodesics with finite cut time. Finally, and as a straightforward consequence of our results, we derive an explicit lower bound for the small time heat kernel asymptotics at the points of C. The question whether C coincides with the cut locus for $k \geq 4$ remains open.

We also edited the two volumes [13] and [12], containing some of the lecture notes of the courses given during the IHP triemster on "Geometry, Analysis and Dynamics on sub-Riemannian Manifolds" which we organized in Fall 2014. The second volume also contains a chapter [11] co-authored by members of the team.

7.2. New results: quantum control

- In recent years, several sufficient conditions for the controllability of the Schrödinger equation have been proposed. In [16], we discuss the genericity of these conditions with respect to the variation of the controlled or the uncontrolled potential. In the case where the Schrödinger equation is set on a domain of dimension one, we improve the results in the literature, removing from the previously known genericity results some unnecessary technical assumptions on the regularity of the potentials.

7.3. New results: neurophysiology

In [4] we propose a supervised object recognition method using new global features and inspired by the model of the human primary visual cortex V1 as the semidiscrete roto-translation group $\mathrm{SE}(2, N) = \mathbb{Z}_N \rtimes \mathbb{R}^2$. The proposed technique is based on generalized Fourier descriptors on the latter group, which are invariant to natural geometric transformations (rotations, translations). These descriptors are then used to feed an SVM classifier. We have tested our method against the COIL-100 image database and the ORL face database, and compared it with other techniques based on traditional descriptors, global and local. The obtained results have shown that our approach looks extremely efficient and stable to noise, in presence of which it outperforms the other techniques it has been compared with.

7.4. New results: switched systems

- In [8] we address the exponential stability of a system of transport equations with intermittent damping on a network of $N \geq 2$ circles intersecting at a single point O. The N equations are coupled through a linear mixing of their values at O, described by a matrix M. The activity of the intermittent damping is determined by persistently exciting signals, all belonging to a fixed class. The main result is that, under suitable hypotheses on M and on the rationality of the ratios between the lengths of the circles, such a system is exponentially stable, uniformly with respect to the persistently exciting signals. The proof relies on a representation formula for the solutions of this system, which allows one to track down the effects of the intermittent damping. A similar representation formula is used in [18] to study the relative controllability of linear difference equations with multiple delays in the state. Thanks to such formula, we characterize relative controllability in time T in terms of an algebraic property of the matrix-valued coefficients, which reduces to the usual Kalman controllability criterion in the case of a single delay. Relative controllability is studied for solutions in the set of all functions and in the function spaces L^p and C^k. We also compare the relative controllability of the system for different delays in terms of their rational dependence structure, proving that relative controllability for some delays implies relative controllability for all delays that are "less rationally dependent" than the original ones. Finally, we provide an upper bound on the minimal controllability time for a system depending only on its dimension and on its largest delay.
- In [9] we address the stability of transport systems and wave propagation on general networks with time-varying parameters. We do so by reformulating these systems as non-autonomous difference equations and by providing a suitable representation of their solutions in terms of their initial conditions and some time-dependent matrix coefficients. This enables us to characterize the asymptotic behavior of solutions in terms of such coefficients. In the case of difference equations with arbitrary switching, we obtain a delay-independent generalization of the well-known criterion for autonomous

systems due to Hale and Silkowski. As a consequence, we show that exponential stability of transport systems and wave propagation on networks is robust with respect to variations of the lengths of the edges of the network preserving their rational dependence structure. This leads to our main result: the wave equation on a network with arbitrarily switching damping at external vertices is exponentially stable if and only if the network is a tree and the damping is bounded away from zero at all external vertices but at most one.

POEMS Project-Team

7. New Results

7.1. New schemes for time-domain simulations

7.1.1. Solving the Homogeneous Isotropic Linear Elastodynamics Equations Using Potentials

Participant: Patrick Joly.

This work is done in collaboration with Sébastien Impériale (EPI M3DISIM) and Jorge Albella from the University of Santiago de Compostela. We consider the numerical solution of 2D elastodynamic equations using the decomposition of the displacement fields into potentials. This appears as a challenge for finite element methods. We address here the particular question of free boundary conditions. A stable (mixed) variational formulation of the evolution problem is proposed based on a clever choice of Lagrange multipliers. This is expected to be efficient when the velocity of shear waves is much smaller than the velocity of pressure waves, since one can adapt the discretization to each type of waves.

7.1.2. Discontinuous Galerkin method with high-order absorbing boundary conditions

Participant: Axel Modave.

This work is done in collaboration with Andreas Atle from TOTAL, Jesse Chan from Rice University and Tim Warburton from Virginia Tech.

Discontinuous Galerkin finite element schemes exhibit attractive features for large-scale time-domain wave-propagation simulations on modern parallel architectures (e.g. GPU clusters). For many applications, these schemes must be coupled with non-reflective boundary treatments to limit the size of the computational domain without losing accuracy or computational efficiency, which remains a challenging task.

We propose a combination of a nodal discontinuous Galerkin method with high-order absorbing boundary conditions (HABCs) for cuboidal computational domains. Compatibility conditions are derived for HABCs intersecting at the edges and the corners of a cuboidal domain. We propose a GPU implementation of the computational procedure, which results in a multidimensional solver with equations to be solved on 0D, 1D, 2D and 3D spatial regions. Numerical results demonstrate both the accuracy and the computational efficiency of our approach.

7.2. Integral equations

7.2.1. Mesh adaptation for the fast multipole method in acoustics

Participants: Faisal Amlani, Stéphanie Chaillat, Samuel Groth.

This work is done in collaboration with Adrien Loseille (EPI Gamma3). We introduce a metric-based anisotropic mesh adaptation strategy for the fast multipole accelerated boundary element method (FM-BEM) applied to exterior boundary value problems of the three-dimensional Helmholtz equation. The present methodology is independent of discretization technique and iteratively constructs meshes refined in size, shape and orientation according to an *optimal* metric reliant on a reconstructed Hessian of the boundary solution. The resulting adaptation is anisotropic in nature and numerical examples demonstrate optimal convergence rates for domains that include geometric singularities such as corners and ridges.

7.2.2. Coupling integral equations and high-frequency methods

Participants: Marc Bonnet, Marc Lenoir, Eric Lunéville, Laure Pesudo, Nicolas Salles.

This theme concerns wave propagation phenomena which involve two different space scales, namely, on the one hand, a medium scale associated with lengths of the same order of magnitude as the wavelength (medium-frequency regime) and on the other hand, a long scale related to lengths which are large compared to the wavelength (high-frequency regime). Integral equation methods are known to be well suited for the former, whereas high-frequency methods such as geometric optics are generally used for the latter. Because of the presence of both scales, both kinds of simulation methods are simultaneously needed but these techniques do not lend themselves easily to coupling.

A first situation, considered by Marc Lenoir, Eric Lunéville and Nicolas Salles, is the scattering of an acoustic wave by two sound-hard obstacles: a large obstacle subject to high-frequency regime relatively to the wavelength and a small one subject to medium-frequency regime. The technique proposed in this case consists in an iterative method which allows to decouple the two obstacles and to use Geometric Optics for the large obstacle and Boundary Element Method for the small obstacle. The method is implemented on the XLife++ library developed in the lab.

The second situation, undertaken in the context of the PhD thesis of Laure Pesudo, is the subject of a partnership with CEA LIST and a collaboration with Francis Collino. Modelling ultrasonic non destructive testing (NDT) experiments simultaneously involves the scattering of waves by defects of moderate size (for which discretization-based methods such as the BEM are appropriate) and propagation over large distances (requiring high-frequency approximations). A new hybrid strategy between the boundary element method (BEM) and ray tracing is proposed in order to allow the accurate and quick simulation of high frequency Non Destructive Testing (NDT) configurations involving diffraction phenomena. Results from its implementation to 2D acoustic NDT-like diffraction configurations have been obtained. The strategy proposed is however generic, and can be extended to three-dimensional configurations and elastodynamic wave propagation.

7.2.3. Dynamic soil-structure interaction

Participants: Marc Bonnet, Stéphanie Chaillat, Zouhair Adnani.

This work, undertaken in the context of the PhD thesis of Zouhair Adnani (CIFRE partnership with EDF), concerns the simulation of dynamic soil-structure interaction (SSI) in connection with seismic assessment of civil engineering structures. Because of the complementary specificities of the finite element method (FEM) and the boundary element method (BEM), it is natural to use the BEM to model the unbounded soil domain, while the FEM is applied for the bounded region comprising the structure undergoing assessment, and possibly its close-range soil environment.

The originality of this work is to formulate, implement, and evaluate on realistic test examples, a computational strategy that combines the fast multipole accelerated boundary element method (visco-elastodynamic COFFEE solver), and the EDF in-house FEM code Code_Aster. In a preliminary phase, the evaluation of transient elastodynamic responses via the Fourier synthesis of frequency domain solutions computed using COFFEE (see Section 5.1) has been studied on several test problems, achieving substantial improvements of computational efficiency for this component of SSI analysis.

The coupling between the two methods is then done in a black-box fashion with the substructuring method by computing the soil impedance (i.e. elastodynamic Poincaré-Steklov) operator relating forces to displacements on the FEM-BEM coupling interface. One of the main challenges is that this operator cannot be assembled due to the iterative nature of the FM-BEM and the potentially large number of degrees of freedom supported by the interface. To reduce the computational costs, we instead compute its projection on a reduced basis of interface modes, which requires to perform as many FM-BEM calculations as interface modes selected. This approach has so far been compared to reference solutions and validated for superficial and buried foundations on homogeneous or heterogeneous soil.

7.2.4. Volume Integral Formulations

Participant: Marc Bonnet.

Volume integral equations (VIEs), also known as Lippmann-Schwinger integral equations, arise naturally when considering the scattering of waves by penetrable, and possibly heterogeneous, inhomogeneities embedded in a homogeneous background medium (for which a fundamental solution is explicitly known). Their derivation and use in e.g. acoustics, elastodynamics or electromagnetism goes back several decades. Since their geometrical support is confined to the spatial region where material properties differ from the background, VIEs are in particular useful for the derivation and justification of homogenized or asymptotic models (the latter providing our main motivation for this study, in connection with [section gradient topologique]). By directly linking remote measurements to unknown inhomogeneities, VIEs also provide a convenient forward modeling approach for medium imaging inverse problems. However, whereas the theory of boundary integral equations is extensively documented, the mathematical properties of VIEs have undergone a comparatively modest coverage, much of it pertaining to electromagnetic scattering problems.

In this work, we investigate the solvability of VIE formulations arising in elastodynamic scattering by penetrable obstacles. The elasticity tensor and mass density are allowed to be smoothly heterogeneous inside the obstacle and may be discontinuous across the background-obstacle interface, the background elastic material being homogeneous. Both materials may be anisotropic, within certain limitations for the background medium.

Towards this goal, we have introduced a modified version of the singular volume integral equation (SVIE) governing the corresponding elastostatic (i.e. zero frequency) problem, and shown it to be of second kind involving a contraction operator, i.e. solvable by Neumann series, for any background material and inhomogeneity material and geometry. Then, the solvability of VIEs for frequency-domain elastodynamic scattering problems follows by a compact perturbation argument, assuming uniqueness to be established. In particular, in an earlier work, we have established a uniqueness result for the anisotropic background case (where, to avoid difficulties associated with existing radiation conditions for anisotropic elastic media, we have proposed an alternative definition of the radiating character of solutions, which is equivalent to the classical Sommerfeld-Kupradze conditions for the isotropic background case). This investigation extends work by Potthast (1999) on 2D electromagnetic problems (transverse-electric polarization conditions) involving orthotropic inhomogeneities in a isotropic background, and contains recent results on the solvability of Eshelby's equivalent inclusion problem as special cases. The proposed modified SVIE is also useful for fixed-point iterative solution methods, as Neumannn series then converge (i) unconditionally for static problems and (ii) on some inhomogeneity configurations for which divergence occurs with the usual SVIE for wave scattering problems.

7.3. Domain decomposition methods

7.3.1. Transparent boundary conditions with overlap in unbounded anisotropic media

Participants: Anne-Sophie Bonnet-Ben Dhia, Sonia Fliss, Yohanes Tjandrawidjaja.

This work is done in the framework of the PhD of Yohanes Tjandrawidjaja, funded by CEA-LIST, in collaboration with Vahan Baronian form CEA. This follows the PhD of Antoine Tonnoir (now Assistant Professor at Insa of Rouen) who developed a new approach, the Half-Space Matching Method, to solve scattering problems in 2D unbounded anisotropic media. The objective is to extend the method to a 3D plate of finite width.

In 2D, our approach consists in coupling several plane-waves representations of the solution in half-spaces surrounding the defect with a FE computation of the solution around the defect. The difficulty is to ensure that all these representations match, in particular in the infinite intersections of the half-spaces. It leads to a Fredholm formulation which couples, via integral operators, the solution in a bounded domain including the defect and some traces of the solution on the edge of the half-planes.

The extension to 3D elastic plates requires some generalizations of the formulation which are not obvious. In particular, we have to use Neumann traces of the solution, which raises difficult theoretical questions.

As a first step, we have considered a scattering problem outside a convex polygonal scatterer for a general class of boundary conditions, using the Half-Space Matching Method. Using the Mellin Transform, we are able to show that this system is coercive + compact in presence of dissipation.

Let us mention that the Half-Space Matching Method has been extended successfully by Julian Ott (Karlsruhe Institut für Technologie) to the scattering by junctions of open waveguides in 2D.

7.3.2. Domain Decomposition Methods for the neutron diffusion equation

Participants: Patrick Ciarlet, Léandre Giret.

This work is done in collaboration with Erell Jamelot (CEA-DEN, Saclay) and Félix Kpadonou (LMV, UVSQ). Studying numerically the steady state of a nuclear core reactor is expensive, in terms of memory storage and computational time. In its simplest form, one must solve a neutron diffusion equation with low-regularity solutions, discretized by mixed finite element techniques, within a loop. Iterating in this loop allows to compute the smallest eigenvalue of the system, which determines the critical, or non-critical, state of the core. This problem fits within the framework of high performance computing so, in order both to optimize the memory storage and to reduce the computational time, one can use a domain decomposition method, which is then implemented on a parallel computer: this is the strategy used for the APOLLO3 neutronics code. The development of non-conforming DD methods for the neutron diffusion equation with low-regularity solutions has recently been finalized, cf. [PC,EJ,FK'1x]. The theory for the eigenvalue problem is also understood. The current research now focuses on the numerical analysis of the full suite of algorithms to prove convergence for the complete multigroup SPN model (which involves coupled diffusion equations).

7.4. Wave propagation in complex media

7.4.1. Perfectly Matched Layers in plasmas and metamaterials

Participants: Eliane Bécache, Maryna Kachanovska.

In this work we consider the problem of the modelling of 2D anisotropic dispersive wave propagation in unbounded domains with the help of perfectly matched layers (PML). We study the Maxwell equations in passive media with the frequency-dependent diagonal tensor of dielectric permittivity and magnetic permeability. An application of the traditional PMLs to this kind of problems often results in instabilities, due to the presence of so-called backward propagating waves. In previous works, this instability was overcome with the help of the frequency-dependent correction of the PML, for isotropic dispersive models.

We show that this idea can be extended to a more general class of models (uniaxial cold plasma, some anisotropic metamaterials). Crucially, we base our considerations on the Laplace-domain techniques. This allows to avoid the analysis of the group and phase velocity (used before) but study (rather formally) coercivity properties of the sesquilinear form corresponding to the PML model in the Laplace domain. The advantage of this method is that it permits to treat problems with dissipation, and provides an intuition on how to obtain explicit energy estimates for the resulting PML models in the time domain. However, such analysis does not allow to obtain easily the necessary stability condition of the PML. We demonstrate that the necessary stability conditions of the PML can be rewritten for a class of models in a form that is easy to verify, and demonstrate that these conditions are sufficient for the stability of the new PMLs with the help of the Laplace-domain techniques. Thanks to the Laplace domain analysis, we are able to rewrite a PML system in the time domain in a form, for which the derivation of energy estimates is simplified (compared to other formulations).

7.4.2. Transparent Boundary Conditions for the Wave Propagation in Fractal Trees

Participants: Patrick Joly, Maryna Kachanovska.

This work, done in collaboration with Adrien Semin (Postdoctoral student at the Technische Universität of Berlin), is dedicated to an efficient resolution of the wave equation in self-similar trees (e.g. wave propagation in a human lung). In this case it is possible to avoid computing the solution at deeper levels of the tree by using the transparent boundary conditions. The corresponding DtN operator is defined by a functional equation in the frequency domain. In this work we propose and compare two approaches to the discretization of this operator in the time domain. The first one is based on the multistep convolution quadrature, while the second one stems from the rational approximations.

7.4.3. High order transmission conditions between homogeneous and homogenized periodic half-spaces

Participants: Sonia Fliss, Valentin Vinoles.

This work is a part of the PhD of Valentin Vinoles, and is done in collaboration with Xavier Claeys from Paris 6 University and EPI ALPINE. It is motivated by the fact that classical homogenization theory poorly takes into account interfaces, which is particularly unfortunate when considering negative materials, because important phenomena arise precisely at their surface (plasmonic waves for instance). To overcome this limitation, we want to construct high order transmission conditions. For now, we have treated the case of a plane interface between a homogeneous and a periodic half spaces. Using matched asymptotic techniques, we have derived high order transmission conditions. We have then introduced an approximate model associated to this asymptotic expansions which consists in replacing the periodic media by an effective one but the transmission conditions are not classical. The obtained conditions involve Laplace- Beltrami operators at the interface and requires to solve cell problems in periodicity cell (as in classical homogenisation) and in infinite strips (to take into account the phenomena near the interface). We establish well posedness for the approximate and error estimate which justify that this new model is more accurate near the interface and in the bulk. From a numerical point of view, the only difficulty comes from the problems set in infinite strips (one half is homogeneous and the other is periodic). This is overcome using DtN operators corresponding to the homogeneous and the periodic media.The numerical results confirm the theoretical ones.

7.4.4. Finite Element Heterogeneous Multiscale Method for Maxwell's Equations

Participants: Patrick Ciarlet, Sonia Fliss.

This work is done in collaboration with Christian Stohrer (Karlsruhe Institute of Technology, Allemagne). In recent years, the Finite Element Heterogeneous Multiscale Method (FE-HMM) has been used to approximate the effective behavior of solutions to PDEs in highly oscillatory media. We started on the extension of the FE-HMM to the Helmholtz equation in such media, and recently we solved the time-harmonic Maxwell equations case. Using a combination of results regarding the FE-HMM and the notion of T-coercivity applied to Maxwell's equations, we derive an a priori error bound and the error. Moreover, numerical experiments corroborate the analytical findings, cf. [PC,SF,CS'1x].

7.5. Spectral theory and modal approaches for waveguides

7.5.1. Plasmonic waveguides

Participants: Anne-Sophie Bonnet-Ben Dhia, Patrick Ciarlet.

This work is done in collaboration with Camille Carvalho (UC Merced, California, USA) and Lucas Chesnel (EPI DEFI). A plasmonic waveguide is a cylindrical structure consisting of metal and dielectric parts. In a certain frequency range, the metal can be seen as a lossless material with a negative dielectric permittivity. The study of the modes of a plasmonic waveguide is then presented as a model eigenvalue problem with a sign-change of coefficients in the main part of the operator. Depending on the values of the contrast of permittivities at the metal/dielectric interface, different situations may occur. We focus on the situation where the interface between metal and dielectric presents corners. For a particular contrast range, the problem is neither self-adjoint nor with compact resolvent, this is the "critical" case. Whereas in the "nice" case, the problem is self-adjoint with compact resolvent and admits two sequences of eigenvalues tending to $-\infty$ and $+\infty$. In the "critical" case, Kondratiev's theory of singularities allows to build extensions of the operator, with compact resolvent. We show that the eigenvalues for one of these extensions can be computed by combining finite elements and Perfectly Matched Layers at the corners. The paradox is that a specific treatment has to be done to capture the corners singularities, even to compute regular eigenmodes. In the "nice" case, we propose and analyze numerical techniques based on the notion of T-coercive meshes that allow to solve the model problem.

7.5.2. Modal analysis of electromagnetic dispersive media

Participants: Christophe Hazard, Sandrine Paolantoni.

We investigate the spectral effects of an interface between a usual dielectric and a negative-index material (NIM), that is, a dispersive material whose electric permittivity and magnetic permeability become negative in some frequency range. We consider here an elementary situation, namely, 1) the simplest existing model of NIM : the Drude model (for which negativity occurs at low frequencies); 2) a two-dimensional scalar model derived from the complete Maxwell's equations; 3) the case of a simple bounded cavity: a camembert-like domain partially

lled with a portion of non dissipative Drude material. Because of the frequency dispersion (the permittivity and permeability depend on the frequency), the spectral analysis of such a cavity is unusual since it yields a nonlinear eigenvalue problem. Thanks to the use of an additional unknown, we show how to linearize the problem and we present a complete description of the spectrum.

7.5.3. Formulation of invisibility in waveguides as an eigenvalue problem

Participants: Antoine Bera, Anne-Sophie Bonnet-Ben Dhia.

This work is done in collaboration with Lucas Chesnel from EPI DEFI, Vincent Pagneux from Laboratoire d'Acoustique de l'Université du Maine and Sergei Nazarov from Russian Academy of Sciences. A scatterer placed in an infinite waveguide may be *invisible* at particular discrete frequencies. We consider two different definitions of invisibility: no reflection (but possible conversion or phase shift in transmission) or perfect invisibility (the scattered field is exponentially decaying at infinity). Our objective is to show that the invisibility frequencies can be characterized as eigenvalues of some spectral problems. Two different approaches will be used for the two different definitions of invisibility, leading to non-selfadjoint eigenvalue problems. Concerning the non-reflection case, our approach based on an original use of PMLs allows to extend to higher dimension and to complex eigenvalues the results obtained by Hernandez-Coronado, Krejcirik and Siegl on a 1D model problem.

7.5.4. Transparent boundary conditions for general waveguide problems

Participants: Anne-Sophie Bonnet-Ben Dhia, Sonia Fliss.

In this work, done in collaboration with Antoine Tonnoir from INSA of Rouen, we propose a construction of transparent boundary conditions which can be used for quite general waveguide problems. Classical Dirichlet-to-Neumann maps used for homogeneous acoustic waveguides can be constructed using separation of variables and the orthogonality of the modes on one transverse section. These properties are also important for the mathematical and numerical analysis of problems involving DtN maps. However this framework does not extend directly to strati

ed, anisotropic or periodic waveguides and for Maxwell's or elastic equations. The difficulties are that (1) the separation of variables is not always possible and (2) the modes of the waveguides are not necessarily orthogonal on the transverse section. We propose an alternative to the DtN maps which uses two artificial boundaries and is constructed using a more general orthogonality property.

7.6. Inverse problems

7.6.1. Linear Sampling Method with realistic data in waveguides

Participants: Laurent Bourgeois, Arnaud Recoquillay.

Our activities in the field of inverse scattering in waveguides with the help of sampling methods has now a quite long history. We now intend to apply these methods in the case of realistic data, that is surface data in the time domain. This is the subject of the PhD of Arnaud Recoquillay. It is motivated by Non Destructive Testing activities for tubular structures and is the object of a partnership with CEA List (Vahan Baronian).

Our strategy consists in transforming the time domain problem into a multi-frequency problem by the Fourier transform. This allows us to take full advantage of the established efficiency of modal frequency-domain sampling methods. We have already proved the feasibility of our approach in the 2D acoustic and 2D elastic case. In particular, we have shown how to optimize the number of sources/receivers and the distance between them in order to obtain the best possible identification result. Experiments are currently carried in CEA.

7.6.2. The "exterior approach" to solve inverse obstacle problems

Participants: Laurent Bourgeois, Arnaud Recoquillay.

We consider some inverse obstacle problems in acoustics by using a single incident wave, either in the frequency or in the time domain. When so few data are available, a Linear Sampling type method cannot be applied. In order to solve those kinds of problem, we propose an "exterior approach", coupling a mixed formulation of quasi-reversibility and a simple level set method. In such iterative approach, for a given defect D, we update the solution u with the help of a mixed formulation of quasi-reversibility while for a given solution u, we update the defect D with the help of a level set method based on a Poisson problem. The case of data in the frequency domain has been studied for the waveguide geometry. We currently investigate the case of data in a finite time domain.

7.6.3. Topological derivatives of leading- and second-order homogenized coefficients.

Participants: Marc Bonnet, Rémi Cornaggia.

This work is done in collaboration with Bojan Guzina from University of Minnesota. We derive the topological derivatives of the homogenized coefficients associated to a periodic material, with respect of the small size of a penetrable inhomogeneity introduced in the unit cell that defines such material. In the context of antiplane elasticity, this work extends existing results to (i) time-harmonic wave equation and (ii) second-order homogenized coefficients, whose contribution reflects the dispersive behavior of the material.

7.6.4. A continuation method for building large invisible obstacles in waveguides

Participants: Antoine Bera, Anne-Sophie Bonnet-Ben Dhia.

In collaboration with Lucas Chesnel (EPI DEFI) and Sergei Nazarov (Saint-Petersburg University), we consider time harmonic acoustic problems in waveguides. We are interested in finding localized perturbations of a straight waveguide which are not detectable in the far field, as they produce neither reflection nor conversion of propagative modes. In other words, such *invisible* perturbation produces a scattered field which is exponentially decaying at infinity in the two infinite outlets of the waveguide.

In our previous contributions, we found a way to build smooth and small perturbations of the boundary which were almost invisible, in the sense that they were producing no reflexions but maybe a phase shift in transmission.

The method is constructive and has been validated numerically. But the drawback is that it is limited to small perturbations. In the present work, we show that the previous idea can be combined with a continuation method, in order to get larger invisible perturbations.

7.7. Aeroacoustics

7.7.1. Time-harmonic acoustic scattering in a vortical flow

Participants: Antoine Bensalah, Patrick Joly, Jean-François Mercier.

This activity is done in the framework of the PhD of Antoine Bensalah, in partnership with Airbus Group. We study the time-harmonic acoustic radiation in a fluid in a general flow which is not curl free, but has restricted vortical areas. The objective is to take into account the complicated coupling between acoustics and hydrodynamics. The Galbrun approach developed previously in 2D is too expensive in terms of degrees of freedom for 3D simulations. As an alternative, we propose to consider instead the Goldstein equations, which are vectorial only in the vortical areas and remain scalar elsewhere.

To begin with, we aim at determining the acoustic field radiated in 2D by a time-harmonic source in a fluid in flow. Goldstein's equations are proved to be well-posed outside a spectrum of frequencies corresponding to resonant streamlines. This band spectrum is explicitly determined for two simple geometries (an annular domain and a rectangular one with periodic conditions). Then the full model is shown to be well-posed under a coercivity condition, implying a subsonic flow with a small enough vorticity.

7.7.2. Propagation of solitons through Helmholtz resonators

Participant: Jean-François Mercier.

With Bruno Lombard (Laboratoire de Mécanique et Acoustique of Marseille), we study the propagation of nonlinear solitary acoustic waves in a 1D waveguide connected to a lattice of Helmholtz resonators. We start from an homogenized model of the literature, consisting of two coupled equations evolution: a nonlinear PDE describing acoustic waves (similar to the Burgers equation), and a linear ODE describing oscillations in the Helmholtz resonators. We have already developed a numerical modeling of this model and we have compared simulations with experimental data.

The drawback of the homogenized model is that all the resonators must be the same. In particular the reflection of an incident wave by a defect cannot be considered. To remedy this limitation, we have proposed an extension of the model, predicting two-way propagation across variable resonators. Thanks to a new discrete description of the resonators, the improved model takes into account two important features: resonators of different strengths and back-scattering effects. Comparisons with experimental data show that a closer agreement is obtained.

SELECT Project-Team

6. New Results

6.1. Model selection in Regression and Classification

Participants: Gilles Celeux, Serge Cohen, Pascal Massart, Sylvain Arlot, Jean-Michel Poggi, Kevin Bleakley.

The well-documented and consistent variable selection procedure in model-based cluster analysis and classification that Cathy Maugis (INSA Toulouse) designed during her PhD thesis in SELECT, makes use of stepwise algorithms which are painfully slow in high dimensions. In order to circumvent this drawback, Gilles Celeux, in collaboration with Mohammed Sedki (Université Paris XI) and Cathy Maugis), have proposed to sort variables using a lasso-like penalization adapted to the Gaussian mixture model context. Using this ranking to select variables, they avoid the combinatory problem of stepwise procedures. The performances on challenging simulated and real data sets are similar to the standard procedure, with a CPU time divided by a factor of more than a hundred.

In collaboration with Jean-Michel Marin (Université de Montpellier) and Olivier Gascuel (LIRMM), Gilles Celeux has continued research aiming to select a short list of models rather a single model. This short list is declared to be compatible with the data using a p-value derived from the Kullback-Leibler distance between the model and the empirical distribution. Furthermore, the Kullback-Leibler distances at hand are estimated through nonparametric and parametric bootstrap procedures. Different strategies are compared through numerical experiments on simulated and real data sets. This year their method has been compared favorably to competing methods.

Sylvain Arlot, in collaboration with Damien Garreau (Inria Paris, Sierra team), studied the kernel change-point algorithm (KCP) proposed by Arlot, Celisse and Harchaoui, that aims at locating an unknown number of change-points in the distribution of a sequence of independent data taking values in an arbitrary set. The change-points are selected by model selection with a penalized kernel empirical criterion. They provide a non-asymptotic result showing that, with high probability, the KCP procedure retrieves the correct number of change-points, provided that the constant in the penalty is well-chosen; in addition, KCP estimates the change-points location at the minimax rate $\log(n)/n$. As a consequence, when using a characteristic kernel, KCP detects all kinds of change in the distribution (not only changes in the mean or the variance), and it is able to do so for complex structured data (not necessarily in $\mathbb{R}^d$). Most of the analysis is conducted assuming that the kernel is bounded; part of the results can be extended when we only assume a finite second-order moment.

Emilie Devijver, Yannig Goude and Jean-Michel Poggi have proposed a new methodology for customer segmentation, in the context of load profiles in energy consumption. The method is based on high-dimensional regression models which perform clustering and model selection at the same time. They have focused on uncovering classes corresponding to different regression models, and compute clustering and model identification in each cluster simultaneously. They have shown the feasibility of the approach on a real data set of Irish customers. Benjamin Goehry is completing a thesis co-supervised by P. Massart and J-M. Poggi, aiming at extending this scheme by introducing the use of time series forecasting models adapted to each cluster.

J-M. Poggi, with J. Cugliari, Y. Goude, have proposed building clustering tools useful for forecasting load consumption. The idea is to disaggregate the global signal in such a way that the sum of disaggregated forecasts significantly improves the prediction of the whole global signal. The strategy has three steps: first they cluster curves defining super-consumers, then they build a hierarchy of partitions from which the best one is selected with respect to a disaggregated forecast criterion. The proposed strategy is applied to a dataset of individual consumers from the French electricity provider EDF.

V. Thouvenot and J-M. Poggi, with A. Pichavant, A. Antoniadis, Y. Goude, consider electricity forecasting using multi-stage estimators of nonlinear additive models. An automatic procedure for variable selection is used to correct middle term forecasting errors for short term forecasting. An application to the EDF customer load demand at an aggregate level is considered as well as an application on load demand from the GEFCom 2012 competition; this is a local application.

6.2. Estimator selection

Participants: Claire Lacour, Pascal Massart.

Estimator selection has become a crucial issue in nonparametric estimation. Two widely used methods are penalized empirical risk minimization (such as penalized log-likelihood estimation) and pairwise comparison (such as Lepski's method). C. Lacour, P. Massart and V. Rivoirard have developed a new method for bandwidth selection which is in some sense intermediate between these two main methods mentioned above, and is called "Penalized Comparison to Overfitting". They have first provided some theoretical results (oracle bounds, minimal penalty) within the framework of kernel density estimation, which leads to some fully data-driven selection strategies. Currently, S. Varet is implementing this method, making a thorough comparison with other selection methods, and tackling the multivariate case. Theoretical work is also in progress, in order to expand the method to other loss functions, such as the Hellinger loss.

6.3. Statistical learning methodology and theory

Participants: Gilles Celeux, Christine Keribin, Michel Prenat, Kaniav Kamary, Sylvain Arlot, Benjamin Auder, Jean-Michel Poggi, Neska El Haouij, Kevin Bleakley.

Gaussian graphical models are widely used to infer and visualize networks of dependencies between continuous variables. However, inferring the graph is difficult when the sample size is small compared to the number of variables. To reduce the number of parameters to estimate in the model, the past PhD. students Emilie Devijver (supervisors: Pascal Massart and Jean-Michel Poggi) and Mélina Gallopin (supervisor: Gilles Celeux) proposed a non-asymptotic model selection procedure supported by strong theoretical guarantees based on an oracle inequality and a minimax lower bound. The covariance matrix of the model is approximated by a block-diagonal matrix. The structure of this matrix is detected by thresholding the sample covariance matrix, where the threshold is selected using the slope heuristic. Based on the block-diagonal structure of the covariance matrix, the estimation problem is divided into several independent problems: subsequently, the network of dependencies between variables is inferred using the graphical lasso algorithm in each block. The performance of the procedure has been illustrated on simulated data. An application to a real gene expression dataset with a limited sample size has been achieved: the dimension reduction allows attention to be objectively focused on interactions among smaller subsets of genes, leading to a more parsimonious and interpretable modular network. This work has been accepted for publication in the *Journal of the American Statistical Association*.

J-M. Poggi, with A. Bar-Hen, have focused on individual observation diagnosis issues for graphical models. The use of an influence measure is a classical diagnostic method to measure the perturbation induced by single elements. The stability issue is here considered using jackknife. For a given graphical model, tools to perform diagnosis on observations are provided. In the second step, a filtering of the dataset to obtain a stable network is proposed.

Latent Block Models (LBM) are a model-based method to cluster simultaneously the d columns and n rows of a data matrix. The Blockcluster package estimates such LBMs. Parameter estimation in LBM is a difficult and multifaceted problem. Although various estimation strategies have been proposed and are now well-understood empirically, theoretical guarantees about their asymptotic behavior is rather rare. Christine Keribin, in collaboration with Mahendra Mariadassou (INRA) and Vincent Brault (Université de Grenoble) have shown that under some mild conditions on the parameter space, and in an asymptotic regime where $\log(d)/n$ and $\log(n)/d$ go to 0 when n and d go to $+\infty$, (1) the maximum likelihood estimate of the complete model (with known labels) is consistent and (2) the log-likelihood ratios are equivalent under the complete and observed (with unknown labels) models. This equivalence allows us to transfer the asymptotic consistency

to the maximum likelihood estimate under the observed model. Moreover, the variational estimator is also consistent. These results extends the results of Bickel et al. (2013) on stochastic block models, and detail the case where the parameter exhibits symmetry.

For the same LBM, Valérie Robert and Yann Vasseur have extended the popular Adjusted Rand Index (ARI) to the task of simultaneous clustering of the rows and columns of a given matrix. This new index, called the Coclustering Adjusted Rand Index (CARI), overcomes the label switching phenomenon while remaining useful and competitive with respect to other indices. Indeed, partitions with high numbers of clusters can be considered, and no convention is required when the numbers of clusters in partitions are different. They are now exploring links with other indices.

Gilles Celeux continued his collaboration with Jean-Patrick Baudry on model-based clustering. This year, they proposed to consider the model selection criterion ICL as a validity index. They show how it can be coupled with a null model of homogeneity focusing on clustering. This null model, which includes the Gaussian distributions, can be difficult to analyze. They find an explicit representation for simple models and show how the parametric bootstrap test can be applied in such situations. In more general situations, they propose a solution for applying this approach involving an "acceptance-rejection" procedure which explores the parameter space to approximate the maximum likelihood estimator inside the null model of homogeneity. The uncovering of this null model highlights the notion of class underlying ICL, and confirms the results of earlier results which show that ICL is consistent for a loss function taking clustering into account.

In collaboration with Arthur White and Jason Wyse (Trinity College, Dublin) Gilles Celeux has evaluated for multivariate Poisson mixture models the performance of a greedy search method compared to the expectation maximization (EM) algorithm, to optimize the ICL model selection criterion, which can be computed exactly for such models. It appears that EM gives often slightly better results, but the greedy search is computationally is more efficient.

The Dutch and French schools of data analysis differ in their approaches to the question: How does one understand and summarize the information contained in a data set? Julie Josse, in collaboration with François Husson (Agro Rennes) and Gibert Saporta (CNAM, Paris), explored the shared factors and differences between the schools, with a focus on methods dedicated to the analysis of categorical data, which are known either as homogeneity analysis (HOMALS) or multiple correspondence analysis (MCA). MCA is a dimension-reduction method which plays a large role in the analysis of tables with categorical nominal variables such as survey data. Though it is usually motivated and derived using geometric considerations, they proved that it amounts to a single proximal Newton step of a natural bilinear exponential family model for categorical data: the multinomial logit bilinear model. They compared and contrasted the behavior of MCA with that of the model on simulations, and discussed new insights into the properties of both exploratory multivariate methods and their cognate models. The main conclusion is to recommend approximating the multilogit model parameters using MCA. Indeed, estimating the parameters of the model is not a trivial task, whereas MCA has the great advantage of being easily solved by a singular value decomposition, as well as being scalable to large datasets.

Julie Josse, with Sobczyk and Bogdan, have discussed the problem of estimating the number of principal components in Principal Components Analysis (PCA). They address this issue by presenting an approximate Bayesian approach based on Laplace approximation, and introduce a general method for building the model selection criteria, called PEnalized SEmi-integrated Likelihood (PESEL). This general framework encompasses a variety of existing approaches based on probabilistic models, like e.g., Bayesian Information Criterion for the Probabilistic PCA (PPCA), and allows for construction of new criteria, depending on the size of the data set at hand. Specifically, they define PESEL when the number of variables substantially exceeds the number of observations. Numerical simulations show that PESEL-based criteria can be quite robust against deviations from probabilistic model assumptions. Selected PESEL-based criteria for estimation of the number of principal components are implemented in the R package varclust, which is available on Github.

Gillies Celeux and Julie Josse have started research on missing data for model-based clustering in collaboration with Christophe Biernacki (Modal, Inria Lille). The aim of this research is to propose appropriate and efficient tools for the packages Mixmod and Mixtcomp.

In collaboration with Jean-Michel Marin (Université de Montpellier) and Christian Robert (Université Paris 9-Dauphine), Gilles Celeux and Kaniav Kamary investigated the ability of Bayesian inference to properly estimate the parameters of Gaussian mixtures in high dimensions. Their study shows how the choice of the prior distributions is important. In particular, independent prior distributions give much better performances. Moreover, when the dimension d becomes very large (say $d > 40$) Bayesian inference becomes questionable. The results of this study will be gathered in a chapter of a book on mixture models that Gilles Celeux is preparing with Christian Robert and Sylvia Fruhwirth Schnatter.

Sylvain Arlot, in collaboration with Robin Genuer (ISPED), studied the reasons why random forests work so well in practice. Focusing on the problem of quantifying the impact of each ingredient of random forests on their performance, they showed that such a quantification is possible for a simple pure forest, leading to conclusions that could apply more generally. Then, they considered "hold-out" random forests, which are a good midpoint between "toy" pure forests and Breiman's original random forests.

J.-M. Poggi and N. El Haouij (with R. Ghozi, S. Sevestre Ghalila and M. Jaïdane) provide a random forest-based method for the selection of physiological functional variables in order to classify stress levels during a real-world driving experience. The contribution of this study is twofold: on the methodological side, it considers physiological signals as functional variables and offers a procedure for data processing and variable selection. On the applied side, the proposed method provides a "blind" procedure of driver's stress level classification that does not depend on expert-based studies of physiological signals.

J-M. Poggi (with R. Genuer, C. Tuleau-Malot, N. Villa-Vialaneix), have focused on random forests in Big Data classification problems, and have performed a review of available proposals about random forests in parallel environments as well as on online random forests. Three variants involving subsampling, Big Data-bootstrap and MapReduce respectively are tested on two massive datasets, one simulated one, and the other, real-world data.

B. Auder and J-M. Poggi (with M. Bobbia, B. Portier) have tested some methods for sequential aggregation for forecasting PM10 concentrations for the next day, in the context of air quality monitoring in Normandy (France). The main originality is that the set of experts contains at the same time statistical models built by means of various methods and groups of predictors, as well as experts coming from deterministic chemical models of prediction. The obtained results show that such a strategy clearly improves the performances of the best expert both in terms of prediction errors and in terms of alerts. What is more, it obtains, for the non-convex weighting strategy, the "unbiasedness" of observed-forecasted scatterplots, which is extremely difficult to obtain.

J-M. Poggi (with A. Antoniadis, I. Gijbels, S. Lambert-Lacroix) have considered the joint estimation and variable selection for mean and dispersion in proper dispersion models. They used recent results on Bregman divergence for establishing theoretical results for the proposed estimators in fairly general settings, and also studied variable selection when there is a large number of covariates, with this number possibly tending to infinity with the sample size. The proposed estimation and selection procedure is investigated via a simulation study, and illustrated via some real data applications.

6.4. Estimation for conditional densities in high dimension

Participants: Claire Lacour, Jeanne Nguyen.

Jeanne Nguyen is working on estimation for conditional densities in high dimension. Much more informative than the regression function, conditional densities are of high interest in recent methods, particularly in the Bayesian framework (studying the posterior distribution). Considering a specific family of kernel estimators, she is studying a greedy algorithm for selecting the bandwidth. Her method addresses several issues: avoiding the curse of high dimensionality under some suitably defined sparsity conditions, being computationally efficient using iterative procedures, and early variable selection, providing theoretical guarantees on the minimax risk.

6.5. Reliability

Participants: Gilles Celeux, Florence Ducros, Patrick Pamphile.

Since June 2015, in the framework of a CIFRE convention with Nexter, Florence Ducros has begun a thesis on the modeling of aging of vehicles, supervised by Gilles Celeux and Patrick Pamphile. This thesis should lead to designing an efficient maintenance strategy according to vehicle use profiles. It involves the estimation of mixtures and competing risk models in a highly-censored setting. Moreover, she can deduce from these models operational tools to estimate the number of spare parts to be stocked in a given period. These tools are defined to take vehicle use patterns into account.

6.6. Statistical analysis of genomic data

Participants: Gilles Celeux, Mélina Gallopin, Christine Keribin, Yann Vasseur, Kevin Bleakley.

The subject of Yann Vasseur's PhD Thesis, supervised by Gilles Celeux and Marie-Laure Martin-Magniette (INRA URGV), is the inference of a regulatory network for Transcriptions Factors (TFs), which are specific genes, of *Arabidopsis thaliana*. For this, a transcriptome dataset with a similar number of TFs and statistical units is available. The first aim consists of reducing the dimension of the network to avoid high-dimensional difficulties. Representing this network with a Gaussian graphical model, the following procedure has been defined:

1. *Selection step*: choose the set of TF regulators (supports) of each TF.
2. *Classification step*: deduce co-factor groups (TFs with similar expression levels) from these supports.

Thus, the reduced network would be built on the co-factor groups. Currently, several selection methods based on Gauss-LASSO and resampling procedures have been applied to the dataset. The study of stability and parameter calibration of these methods is in progress. The TFs are clustered with the Latent Block Model into a number of co-factor groups, selected with BIC or the exact ICL criterion. Since these models are built in an ad hoc way, Yann Vasseur has defined complex simulation tools to asses their performances in a proper way.

In a collaboration with Marie-Laure Martin-Magniette, Cathy Maugis and Andrea Rau, Gilles Celeux has studied gene expression obtained from high-throughput sequencing technology. The focus is on the question of clustering gene expression profiles as a means to discover groups of co-expressed genes. A Poisson mixture model is proposed, using a rigorous framework for parameter estimation, as well as for the choice of the appropriate number of clusters. They illustrate co-expression analyses using this approach on two real RNA-seq datasets. A set of simulation studies also compares the performance of the proposed model with that of several related approaches developed to cluster RNA-seq and serial analysis of gene expression data. The proposed method is implemented in the open-source R package `HTSCluster`, available on CRAN. It can now be compared with Gaussian mixtures obtained after relevant data transformations. Moreover, the performance of `HTSCluster` is compared with $k$means-like algorithms using the χ^2 distance.

In collaboration with Benno Schwikowski, Iryna Nikolayeva and A Anavaj Sakuntabhai (Pasteur Institute, Paris), Kevin Bleakley works on using 2-d isotonic regression to predict dengue fever severity at hospital arrival using high-dimensional microarray gene expression data. Important marker genes for dengue severity have been detected, some of which now have been validated in external lab trials.

6.7. Model based-clustering for pharmacovigilance data

Participants: Gilles Celeux, Christine Keribin, Valérie Robert.

In collaboration with Pascale Tubert-Bitter, Ismael Ahmed and Mohamed Sedki, Gilles Celeux and Christine Keribin have started research concerning the detection of associations between drugs and adverse events in the framework of the PhD of Valerie Robert. At first, this team developed model-based clustering inspired by latent block models, which consists of co-clustering rows and columns of two binary tables, imposing the same row ranking. This enables it to highlight subgroups of individuals sharing the same drug profile, and subgroups of adverse effects and drugs with strong interactions. Furthermore, some sufficient conditions are provided to obtain identifiability of the model, and some results are shown for simulated data. The exact ICL criterion has been extended to this double block latent model. Through computer experiments, Valérie Robert has demonstrated the interest of the proposed model, compared with standard contingency table analysis, to detect co-prescription and masking effects.

6.8. Statistical rating and ranking of scientific journals

Participants: Gilles Celeux, Julie Josse, Simon Grah.

In collaboration with Jean-Louis Foulley (université of Montpellier), Gilles Celeux and Julie Josse have started research on statistical rating and ranking of scientific journals. This research was the subject of the internship of Simon Grah (Université Paris-Sud). Simon Grah compared many models on a set of 47 statistical journals. His study showed that the Row-Column (RC) models appears to be the most relevant. In the future, Bayesian inference for different approaches, including PageRank, will be considered.

TAO Project-Team

7. New Results

7.1. Optimal Decision Making under Uncertainty

The Tao UCT-SIG is working mainly on mathematical programming tools useful for power systems. In particular, we advocate a data science approach, in order to reduce the model error - which is much more critical than the optimization error, in most cases. Real data are the best way for handling uncertainties. Our main results in 2016 are as follows:

- **Noisy optimization** In the context of stochastic uncertainties, noisy optimization handles the model error by simulation-based optimization. Our results include:
  - It has been conjectured that gradient approximation by finite differences (hence, not a comparison-based method) is necessary for reaching such a simple regret of O(1/N). We answer this conjecture in the negative [32], providing a comparison-based algorithm as good as gradient methods, i.e. reaching O(1/N) - under the condition, however, that the noise is Gaussian.
  - The concept of Regret is widely used in the bandit literature for assessing the performance of an algorithm. The same concept is also used in the framework of optimization algorithms, sometimes under other names or without a specific name. Experimental results on the noisy sphere function show that the approximation of Simple Regret, termed Approximate Simple Regret, used in some optimization testbeds, fails to estimate the Simple Regret convergence rate, and propose a new approximation of Simple Regret, the Robust Simple Regret [22].
- **Capacity Expansion Planning** The optimization of capacities in large scale power systems is a stochastic problem, because the need for storage and connections (i.e. exchange capacities) varies a lot from one week/season to another. It is usually tackled through sample average approximation, i.e. assuming that the system which is optimal on average over the last 40 years (corrected for climate change) is also approximately optimal in general. However, in many cases, data are high-dimensional; the sample complexity, i.e. the amount of data necessary for a relevant optimization of capacities, increases linearly with the number of parameters and can be scarcely available at the relevant scale. This leads to an underestimation of capacities. We suggested the use of bias correction in capacity estimation, and investigated the importance of the bias phenomenon, and the efficiency of both standard and original bias correction tools [53].
- **Multi-armed bandits** We studied the problem of sequential decision making in the context of multi-armed bandits. We provided:
  - An algorithm to handle a non-stationary formulation of the stochastic multi-armed bandit where the rewards are not assumed to be identically distributed, that achieves both a competitive regret and sampling complexity against a best sequence of arms. See [61].
  - An algorithm to handle the task of recommending items (actions) to users sequentially interacting with a recommender system. Users are modeled as latent mixtures of C many representative user classes, where each class specifies a mean reward profile across actions. Both the user features (mixture distribution over classes) and the item features (mean reward vector per class) are unknown a priori. The user identity is the only contextual information available to the learner while interacting. This induces a low-rank structure on the matrix of expected rewards from recommending item a to user b. The problem reduces to the well-known linear bandit when either user-or item-side features are perfectly known. In the setting where each user, with its stochastically sampled taste profile, interacts only for a small number of sessions, we develop a bandit algorithm for the two-sided uncertainty. It combines the Robust Tensor Power Method with the OFUL linear bandit algorithm. We provide the first rigorous regret analysis of this combination. See [63].

- **Confidence intervals for streaming data** We consider, in a generic streaming regression setting, the problem of building a confidence interval (and distribution) on the next observation based on past observed data. The observations may have arbitrary dependency on the past observations and come from some external filtering process making the number of observations itself a random stopping time. In this challenging context, we provide confidence intervals based on self-normalized vector-valued martingale techniques, applied to the estimation of the mean and of the variance. See [69].
- **Forecasting tool for Hydraulic networks** We studied a problem of prediction in the context of the monitoring of an hydraulic network by the French company Prolog-ingenierie. The problem is to predict the value of some specific sensor in the next thirty minutes from the activity of the network (values of all other sensors) in the recent past. We designed a simple tool for that purpose, based on a random forests. The tool has been tested on data generated from the activity recorded on the Parisian hydraulic network in 2010, 2011 and 2013.

7.2. Continuous Optimization

- **Markov Chain Analysis of Evolution Strategies** The theory of Markov chains with discrete time and continuous state space turns out to be very useful to analyze the convergence of adaptive evolution strategies, including simplified versions of the state-of-the art CMA-ES. Exploiting invariance properties of the objective function and of a wide variety of comparison-based optimisation algorithms, we have developed a general methodology to prove global linear convergence [4]. The constructed Markov chains also show the connection between comparison-based adaptive stochastic algorithms and Markov chain Monte Carlo algorithms. Furthermore, we have continued to work on new theoretical tools that exploit deterministic control models to prove the irreducibility and T-chain property of general Markov chains. These tools promise to trivialise some stability proofs of the Markov chains we are interested in to analyse.
- **Large-scale Optimisation Algorithms** We have been working on (improved) variants of CMA-ES with more favorable scaling properties with the dimension. While computing and using the natural gradient in appropriate subspaces turned out to be considerably more difficult than expected, we explored variants that restrict the covariance via projection, so-called VkD-CMA-ES [21]. We derived a computationally efficient way to update the restricted covariance matrix, where the richness of the model is controlled by the integer parameter k. This parameter provides a smooth transition between the case where only diagonal elements are subject to changes and changes of the full covariance matrix. In the latter case, the update is equivalent with the original CMA-ES. In order to get rid of the control parameter we propose an adaptation of k which turns out to be surprisingly efficient [20].
- **Analysis of Lagrangian based Constraints Handling in Evolution Strategies** We have addressed the question of linear convergence of evolution strategies on constrained optimisation problems with one linear constraint. Based on previous works, we consider an adaptive augmented Lagrangian approach for the simple (1+1)-ES [23] and for the CMA-ES [24]. By design both algorithms derive from a framework with an underlying homogenous Markov chain which paves the way to prove linear convergence on a comparatively large class of functions. For the time being, stability of the Markov chain, associated with linear convergence, has been shown empirically on convex-quadratic and ill-conditioned functions.
- **Benchmarking of continuous optimizers** We have been pursuing our efforts towards improving the standards in benchmarking of continuous optimisers [65], [66], [64]. Three new testbeds have been developed and implemented. (i) A bi-objective testbed [74] where also a corresponding performance assessment procedure has been advised [62]. In this context, a new version of MO-CMA-ES has been developed and benchmarked [44] on this testbed. (ii) A large-scale testbed, as a straight forward extension of the standard tested. The extension is based on a general methodology we have developed to construct non-trivial but scalable test functions [19]. (iii) a constrained testbed (unpublished).

7.3. Data Science

- **High Energy Physics** The focus of the period has been to expand the collaboration with the High Energy Physics experiments started with the success of the 2014 HiggsML challenge [18] to new issues. The subject of V. Estrade Phd is to advance domain adaptation methods in the specific context of uncertainty quantification and calibration. So far, transfer learning has been addressed only with classical, additive and differentiable objective functions as performance criteria. However, learning to discover, exemplified by HEP, relies on more global and difficult criteria, related to the Area Under Roc Curve (AUC) and Neymann-Pearson learning. CERN funds another PhD (A. Pol), on anomaly detection. Another promising theme has emerged with the ongoing organization of a Tracking Challenge (TrackML) [56], [72], which focuses on extreme scaling of ML image processing.
- **Personal Semantics** Our algorithm for inducing a taxonomy from a set of domain terms, that was placed first in the international Taxonomy Induction task, part of the SemEval 2015 conference in Denver, has been improved by the development of a robust technique for discovering the domain vocabulary for a new topic using a directed crawler we created. We have created hundreds of taxonomy for personal themes (hobbies, illnesses) that can be integrated into our Personal Semantics platform PTraces, and have deployed and evaluated the taxonomies. We also have introduced newer machine learning methods, such as Latent Dirichlet Allocation, for better recognition of domain vocabularies [55], [71].
- **Distributed system observation** The work on distributed system automated analaysis and description has been persued thru the continued development of the GAMA multi-agent framework https://github.com/gama-platform/gama/wiki. The simulation framework has been applied to the study of a new protocol for MOOC management [6]. Philippe Caillou is associated to the young researcher ANR ACTEUR, coordinated by Patrick Taillandier (IDEES, Rouen university). With this project, the BDI cognitive agent model has been improved both in term of flexibility and ease of use for the non expert modeler [50].
- **Computational social sciences** Thomas Schmitt's PhD focuses on the matching of job offers and applicant CVs. An informal collaboration with the Qapa agency (FUI proposal underway) gave us access to the 2012-2016 logs of their activity (CVs, job announcements and application clicks). This wealth of data delivered some unexpected findings, e.g., as to the differences between people's practice (the clicks) and their say (the documents). In [49], with Philippe Caillou and Michèle Sebag, a deep NN system MAJORE (MAtching JObs and REsumes) was proposed, trained to match the metric properties extracted from the collaborative filtering matrix, and address the cold start problem. A further research perspective, in collaboration with J.-P. Nadal from EHESS, is to build an observatory of the job demand dynamics.

  The Cartolabe project, started in Feb. 2016 (F. Louistisserand's engineer stint), applies machine learning techniques to build an interpretable representation from vast amounts of scientific articles. The goal is to use raw textual data, and the results of the pre-processing chain achieved by ANHALYTICS, to define a topology on authors, scientific themes, and teams, and enforce its 2D projection in a semantically admissible way. The collaboration with AVIZ is key to enable the scalable and navigable exploitation of this map. The perspective for 2017 is to build a visual interrogation of the map (locating all author names relevant to a given request) and to display the temporal evolution of the research activities.
  Amiqap studies the relation between quality of life at work and company performance, using both survey data on individual workers (collected by DARES, the statistical service of the French Ministry of labor, in 2013) and administrative data on companies provided by SECAFI, a union body. The study is run by a team within TAO (Philippe Caillou, Isabelle Guyon, Michèle Sebag and Paola Tubaro, plus post-doctoral researcher Olivier Goudet and intern Diviyan Kalainathan) in collaboration with Mines ParisTech social science and economics (SES) department, the RITM economics research center (Univ. Paris Sud) and the think-tank La Fabrique de l'Industrie. In its first stage, the exploratory analysis delivered some unexpected results, e.g. as to the existence of a "industry worker cluster", or the non-monotonous relationship between autonomy, salary and subjective satisfaction. A summary of these findings has been released online on the website of La

Fabrique de l'Industrie, as a complement to their book on the same topic (published in October 2016). The exploratory analysis of the SECAFI data (yet unpublished) complements the above and shows how workers' satisfaction correlates with companies' financial and social performance indicators, though with marked differences across industries. The key question regards the nature of this relationship: cause, effect or due to a confounder feature (the industrial sector). Further research (Diviyan Kalainathan's PhD, O. Goudet post-doc) will focus on the use and extension of causal modelling algorithms on this issue; these perspectives attract quite some interest from the ministry (DARES) and big industrial players, willing to assess the relevance of their HR policies.

7.4. Designing criteria

- **Algorithm selection and configuration** Two PhD theses are related to the former *Crossing the Chasm* SIG: Nacim Belkhir (CIFRE PhD with Thalès) is working on Per Instance Algorithm Configuration (PIAC) in the context of continuous optimization. He has worked on the use of surrogate models for feature computation in case of expensive objective functions [31] and has validated his work with Differential Evolution applied to BBOB testnehc [30]. Defence planned for March 2017.

  François Gonard's PhD is dedicated to optimization algorithm selection. The original application domain was that of expensive car industry simulations (within the IRT-ROM project). The lack of real test cases made him investigate some combinatorial optimization setting, for which there exist public datasets. François obtained a "Honorable mention from the jury" for his submission to the ICON Challenge (http://iconchallenge.insight-centre.org/), for its original approach coupling a pre-scheduler and an algorithm selector [39]. Defence is planned for November 2017.

  The work done during Mustafa Misir's post-doc stint (ERCIM 2013-2014), regarding the formalization and tackling of the algorithm selection problem in terms of a collaborative filtering problem, was finally published [15].
- **A statistical physics perspective** Our activity on probabilistic model design is progressively moving from static explicit interactions to dynamical ones and to latent variable models, taking inspiration from latent feature representations provided by deep learning techniques. Concerning explicit pairwise interactions models like in [14] initially motivated by traffic applications, a systematic treatment of loop corrections based on a minimal cycle basis [11] has led us to propose: (i) a fast and large scale generalized belief propagation method (GCBP) with more robust convergence properties than bare belief propagation (ii) an inverse approximate MRF with linear scaling of the computational time, compliant with GCBP (iii) a new sampling method based on extracting random sub-graph of tree-width 2 on which GCBP can provide exact marginals. More generally considering effect of problematic i.e. frustrated cycles open the possibility for new criteria in model design. In particular we have started to bridge this work with the analysis of multi-layer restricted Boltzmann machines (RBM). Remarkably these possess a planar dual representation and we are expecting the density of frustrated cycles nodes to play a key role when characterizing an RBM learned from structured data by contrast with purely random instances. Additionally we have identify some properties of the data themeselves that have to be taken into consideration when learning static [9] or dynamical [8] Ising models.
- **Artificial Immune Systems** Within the E-Lucid project with Thalès TERESIS, around anomaly detection in network trafic, a first approach has been developed using Artificial Immune System (AIS) and the concept of Voronoi representation. A first proof of concept was a poster at the GECCO conference [70], before a complete paper was published at the PPSN conference [46]. Note that this work on anomaly detection is on-going using Deep Learning. AIS are also the basis of Chaouki Boufenar's PhD work (visiting TAO from U. Oran, Algérie), with a first work on arabic characters recognition [5].

7.5. Deep Learning and Information Theory

- **Neural networks for computer vision** We continued working on the topic of large-scale image segmentation with multiple object detection. The application target is the analysis of high-resolution multispectral satellite images covering the Earth. Challenges are numerous: finding good features to distinguish objects, obtaining fine-resolution segmentations, while dealing with badly-registered groundtruth, keeping a scalable complexity, while avoiding boundary effects when tiling a big image into small ones, which are processed independently and merged back together. We propose to move to fully convolutional neural networks [45] to avoid artifacts from patch-based approaches. We show the benefits of training first on imprecise groundtruth, which is available in large amounts, and then refining on precise but scarce groundtruth [13]. To further refine the segmentation, as convolutional networks tend to produce blurry outputs, we use recurrent neural networks to learn the partial differential equation (PDE) which would sharpen the segmentations, i.e. an iterative process taking into account the edges in the original image to locate precisely their boundaries and to sharpen them [67]. Finally, to benefit simultaneously from information at various resolutions, we design a new, more suitable architecture [68].
  We also started to work on medical image classification, in the long-term goal of automatic diagnosis, in collaboration with the Necker Hospital and the Inria start-up Therapixel, and on image labelling and representation, with the database editor company Armadillo, through the Adamme project (cf Section 9.2.1).
  In collaboration with the University of Barcelona, we organize a series of challenges in video analysis of human behavior (ChaLearn Looking at People series). Looking at People (LAP) is an area of research that deals with the problem of automatically recognizing people in images, detecting and describing body parts, inferring their spatial configuration, performing action/gesture recognition from still images or image sequences, often including multi-modal data. Any scenario where the visual or multi-modal analysis of people takes a key role is of interest to us within the field of Looking at People. We have been leaders in organizing challenges in this area since 2013 [10], [12], [36], organizing events sponsored by DARPA, NSF, Microsoft, Google, Facebook, NVIDIA, and others. In 2016 we organized follow up competitions on gesture recognition [52] and face aging [37] to advance the state-of-the-art in areas we had previously explored. We also organized two rounds of a completely new recognition on personality trait evaluation from short video clips [47], [34]. The purpose of this study is to evaluate whether human first impression judgements are consistent and reproducible. Such research could lead to device coaching curricula to help job applicants present themselved better and hiring managers to overcome unsubstantiated negative biases. The winners of the challenge used Deep Learning methods. The third place winners teamed up with the organizers to put together a demonstration system, which was shown at the NIPS conference(https://nips.cc/Conferences/2016/Schedule?showEvent=6314). Work performed in collaboration with UC Berkeley on fingerprint verification using Deep Learning was also presented in this demonstration.
- **Natural Gradients for Deep Learning** Deep learning is now established as a state-of-the-art technology for performing different tasks such as image or sequence processing. Nevertheless, much of the computational burden is spent on tuning the hyper-parameters. On-going work, started during the TIMCO project, is proposing, in the framework of Riemannian gradient descents, invariant algorithms for training neural networks that effectively reduce the number of arbitrary choices, e.g., affine transformations of the activation functions or shuffling of the inputs. Moreover, the Riemannian gradient descent algorithms perform as well as the state-of-the-art optimizers for neural networks, and are even faster for training complex models. The proposed approach is based on Amari's theory of information geometry and consists in practical and well-grounded approximations for computing the Fisher metric. The scope of this framework, going beyond Deep Learning, encompasses any class of statistical models. This year's contribution is a new, simple framework (both theoretical and practical) that allowed us to release a simpler implementation of these techniques in Torch (one of the main deep learning libraries in use) and demonstrate good performance on real data. We have also started to explore criteria from information geometry criteria for automating the construction and selection of network architectures themselves, a major problem

given the current trend towards highly complex, hand-built model architectures (P. Wolinski's PhD).

- **Training dynamical systems online without backtracking** with application to recurrent neural networks. The standard way to train recurrent neural networks and other systems that exhibit a temporal dynamical behavior involves "backpropagation through time", which as the name indicates goes backward in time and is unrealistic. Last year we proposed an algorithm to learn the parameters of a dynamical system in an online, memoryless setting, thus scalable and requiring no backpropagation through time, in a way guaranteed to be unbiased. This year we started to provide full convergence proofs for this algorithm (the first of their kind). Moreover Corentin Tallec (PhD) proposed a considerably simpler version of the algorithm keeping the same key mathematical properties, which now allows for a simple "black-box" implementation on top of any existing recurrent network model.

TROPICAL Team

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, Antoine Hochart.

The PhD work of Antoine Hochart [12] deals with the applications of methods of non-linear fixed point theory to zero-sum games.

A highlight of his PhD is the characterization of the property of ergodicity for zero-sum games. In the special "zero-player" case, i.e., for a Markov chain equipped with an additive functional (payment) of the trajectory, the ergodicity condition entails that the mean payoff is independent of the initial state, for any choice of the payment. In the case of finite Markov chains, ergodicity admits several characterizations, including a combinatorial one (the uniqueness of the final class). This carries over to the two player case: ergodicity is now characterized by the absence of certain pairs of conjugate invariant sets (dominions), and it can be checked using directed hypergraphs algorithms. This leads to an explicit combinatorial sufficient condition for the solvability of the "ergodic equation", which is the main tool in the numerical approach of the mean payoff problem. These results appeared in [59], [58], [60]. A more general approach was developed in [30], in which zero-sum games are now studied abstractly in terms of accretive operators. This allows one to show that the bias vector (the solution of the ergodic equation) is unique for a generic perturbation of the payments.

Another series of results of the thesis concern the finite action space, showing that the set of payments for which the bias vector is not unique coincides with the union of lower dimensional cells of a polyhedral complex, which an application to perturbation schemes in policy iteration [47].

A last result of the thesis is a representation theorem for "payment free" Shapley operators, showing that these are characterized by monotonicity and homogeneity axioms [48]. This extends to the two-player case known representation theorems for risk measures.

7.1.2. Probabilistic and max-plus approximation of Hamilton-Jacobi-Bellman equations

Participants: Marianne Akian, Eric Fodjo.

The PhD thesis of Eric Fodjo concerns stochastic control problems obtained in particular in the modelisation of portfolio selection with transaction costs. The dynamic programming method leads to a Hamilton-Jacobi-Bellman partial differential equation, on a space with a dimension at least equal to the number of risky assets. The curse of dimensionality does not allow one to solve numerically these equations for a large dimension (greater to 5). We propose to tackle these problems with numerical methods combining policy iterations, probabilistic discretisations, max-plus discretisations, in order to increase the possible dimension. Another solution is to replace policy iterations by an approximation with optimal switching problems.

In [27], [26] (also presented in [35], [23]), 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. We construct a lower complexity probabilistic numerical algorithm by combining the idempotent expansion properties obtained by McEneaney, Kaise and Han [93], [99] for solving such problems with a numerical probabilistic method such as the one proposed by Fahim, Touzi and Warin [78] for solving some fully nonlinear parabolic partial differential equations, when the volatility does not oscillate too much. Numerical tests on a small example of pricing and hedging an option are presented. Moreover, more recently, we improved the method of Fahim, Touzi and Warin to allow one to solve fully nonlinear parabolic partial differential equations with general volatilities.

7.2. Non-linear Perron-Frobenius theory, nonexpansive mappings and metric geometry

7.2.1. Isometries of the Hilbert geometry

Participant: Cormac Walsh.

In a collaboration with Bas Lemmens (Kent University, UK), we have been studying the Hilbert geometry in finite dimensions, especially its horofunction boundary and isometry group. The book chapter [117] contains a survey of this work. However, the infinite dimensional case is also interesting, and has been used as a tool for many years in non-linear analysis. Despite this, very little is known about the geometry of these spaces when the dimension is infinite.

An example of a problem in which we are interested is the following. In finite dimension it is known that a Hilbert geometry is isometric to a normed space if and only if it is a simplex. We have shown [118] that, more generally, a Hilbert geometry is isometric to a Banach space if and only if it is the cross-section of a positive cone, that is, the cone of positive continuous functions on some compact topological space. To solve this problem we found it useful to study the horofunction boundary in the infinite-dimensional case.

We are continuing to study similar problems in relation to this topic in collaboration with Bas Lemmens of the University of Kent.

7.2.2. Volume growth in the Hilbert geometry

Participant: Cormac Walsh.

In a collaboration with Constantin Vernicos of Université Montpellier 2, we are investigating how the volume of a ball in a Hilbert geometry grows as its radius increases. Specifically, we are studying the volume entropy

$$\lim_{r\to\infty} \frac{\log \operatorname{Vol} B(x,r)}{r}, \tag{1}$$

where $B(x,r)$ is the ball with center x and radius r, and Vol denotes some notion of volume, for example, the Holmes–Thompson or Busemann definitions. Note that the entropy does not depend on the particular choice of x, nor on the choice of the volume. It is known that the hyperbolic space, or indeed any Hilbert geometry with a C^2-smooth boundary of strictly positive curvature, has entropy $n-1$, where n is the dimension, and it has recently been proved that this is the maximal entropy possible for Hilbert geometries of the given dimension.

Constantin Vernicos has shown that, in dimension 2 and 3, the volume entropy of a Hilbert geometry on a convex body is equal to exactly twice the *approximability* of the body, that is, the power of $1/\epsilon$ governing the growth of the number of vertices needed to approximate the body by a polytope within ϵ, as ϵ decreases.

Studying polytopal Hilbert geometries, we have demonstrated [53] a close relation between the volume and the number of *flags* of the polytope, more precisely, that the volume of large balls is asymptotically proportional to the number of flags. This suggested to us defining a new notion of approximability using flags rather than vertices. We have shown [53] that the volume entropy of a Hilbert geometry on a convex body is equal to exactly twice this *flag-approximability* in all dimensions. This implies in particular that the volume entropy of a convex body is equal to that of its dual.

7.2.3. The set of minimal upper bounds of two matrices in the Loewner order

Participant: Nikolas Stott.

A classical theorem of Kadison shows that the space of symmetric matrices equipped with the Loewner order is an anti-lattice, meaning that two matrices have a least upper bound if and only if they are comparable. In [52], we refined this theorem by characterizing the set of minimal upper bounds: we showed that it is homeomorphic to the quotient space $O(p) \setminus O(p,q)/O(q)$, where $O(p,q)$ denotes the orthogonal group associated to the quadratic form with signature (p,q), and $O(p)$ denotes the standard pth orthogonal group.

7.2.4. Checking the strict positivity of Kraus maps is NP-hard

Participant: Stéphane Gaubert.

In collaboration with Zheng Qu (now with HKU, Hong Kong), I studied several decision problems arising from the spectral theory of Kraus maps (trace preserving completely positive maps), acting on the cone of positive semidefinite matrices. The latter appear in quantum information. We showed that checking the irreducibility (absence of non-trivial invariant face of the cone) and primitivity properties (requiring the iterates of the map to send the cone to its interior) can be checked in polynomial time, whereas checking positivity (whether the map sends the cone to its interior) is NP-hard. In [17], we studied complexity issues related to Kraus maps, and showed in particular that checking whether a Kraus map sends the cone to its interior is NP-hard.

7.3. Tropical algebra and convex geometry

7.3.1. Formalizing convex polyhedra in Coq

Participants: Xavier Allamigeon, Ricardo Katz [Conicet, Argentine].

We formalize a certain fragment of the theory of convex polyhedra and their combinatorial properties. Our motivation is that convex polyhedra are involved in a wide range of analysis techniques such as in formal verification, and that their combinatorial properties are used to establish more fundamental results, especially in tropical geometry.

This formalization has been conducted in Coq using the Mathematical Components library. We have implemented a full formalization of the simplex algorithm, which allows to make several key properties of convex polyhedra (feasibility, unboundedness, etc) decidable. From this, we have deduced a formal proof of strong duality theorem in linear programming, and of Farkas lemma. We also have a formal implementation of Motzkin's double description method, which provides a constructive way to prove Minkowski theorem for polyhedra.

7.3.2. Tropical totally positive matrices

Participants: Stéphane Gaubert, Adi Niv.

In [50], we investigate 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 show that tropical totally positive matrices essentially coincide with the Monge matrices (defined by the positivity of 2×2 tropical minors), arising in optimal transport. These results have been presented in [41], [40].

7.3.3. Tropical compound matrix identities

Participants: Marianne Akian, Stéphane Gaubert, Adi Niv.

In [55], [57], we proved some identities on matrices using a weak and a strong transfer principles. In the present work, we prove identities on compound matrices in extended tropical semirings. Such identities include analogues to properties of conjugate matrices, powers of matrices and $\mathrm{adj}\,(A)\det(A)^{-1}$, all of which have implications on the eigenvalues of the corresponding matrices. A tropical Sylvester-Franke identity is provided as well. Even though part of these identities hold over any commutative ring, they cannot be adjusted to semirings with symmetry using the existing weak and strong transfer principles. Here, we provide the proofs by means of graph theory arguments.

7.3.4. Supertropical algebra

Participant: Adi Niv.

Several properties of matrices over the tropical algebra are studied using the supertropical algebra introduced in [92].

The only invertible matrices in tropical algebra are diagonal matrices, permutation matrices and their products. However, the pseudo-inverse A^{∇}, defined as $\frac{1}{\det(A)}\operatorname{adj}(A)$, with $\det(A)$ being the tropical permanent, inherits some classical algebraic properties and has some surprising new ones. In [104], defining B and B' to be tropically similar if $B' = A^{\nabla}BA$, we examine the characteristic (max-)polynomials of tropically similar matrices as well as those of pseudo-inverses. Other miscellaneous results include a new proof of the identity for $\det(AB)$ and a connection to stabilization of the powers of definite matrices.

In a joint work with Louis Rowen (Bar Ilan Univ.) [21], we study the pathology that causes tropical eigenspaces of distinct supertropical eigenvalues of a non-singular matrix A, to be dependent. We show that in lower dimensions the eigenvectors of distinct eigenvalues are independent, as desired. The index set that differentiates between subsequent essential monomials of the characteristic polynomial, yields an eigenvalue λ, and corresponds to the columns of the eigenmatrix $A + \lambda I$ from which the eigenvectors are taken. We ascertain the cause for failure in higher dimensions, and prove that independence of the eigenvectors is recovered in case the "difference criterion" holds, defined in terms of disjoint differences between index sets of subsequent coefficients. We conclude by considering the eigenvectors of the matrix $A^{\nabla} := \frac{1}{\det(A)}\operatorname{adj}(A)$ and the connection of the independence question to generalized eigenvectors.

7.3.5. Volume and integer points of tropical polytopes

Participants: Marie Maccaig, Stéphane Gaubert.

We investigated 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 $2^{\text{poly}(m,n)}$ existing. We further proved the ♯P-hardness for the analogous problems for tropical polytopes instead defined by inequalities. We also investigated the relation between the set of integer points of a tropical polytope and the image by the valuation of polytopes over the nonarchimedean field of Puiseux series.

7.3.6. Primal dual pair of max-algebraic integer linear programs (MLP)

Participant: Marie Maccaig.

There are known weak and strong duality theorems for max-algebraic linear programs. I investigated the integer versions of these problems; considering the impact of requiring integer solutions instead of real solutions. I proved a tight bound on the duality gap for a pair of integer solutions to the primal and dual MLPs, and searched for conditions on when the optimal values of the integer primal and dual MLPs coincide.

7.3.7. Tropical Jacobi identity

Participants: Marie Maccaig, Adi Niv.

In a joint work with Sergei Sergeev (Birmingham), we investigated the combinatorial interpretation for the Tropical Jacobi identity. Inspired by Butkovic's paper, "Max-algebra, the algebra of combinatorics?" and many other links between max-algebra and combinatorics, we try to link this tropical quantity to a new type of multiple assignment problem.

7.4. Tropical methods applied to optimization, perturbation theory and matrix analysis

7.4.1. Majorization inequalities for valuations of eigenvalues using tropical algebra

Participants: Marianne Akian, Stéphane Gaubert.

We consider a matrix with entries over the field of Puiseux series, equipped with its non-archimedean valuation (the leading exponent). In [13], with Ravindra Bapat (Univ. New Delhi), we establish majorization inequalities relating the sequence of the valuations of the eigenvalues of a matrix with the tropical eigenvalues of its valuation matrix (the latter is obtained by taking the valuation entrywise). We also show that, generically in the leading coefficients of the Puiseux series, the precise asymptotics of eigenvalues, eigenvectors and condition numbers can be determined. For this, we apply diagonal scalings constructed from the dual variables of a parametric optimal assignment constructed from the valuation matrix.

In recent works with Andrea Marchesini and Françoise Tisseur (Manchester University), we use the same technique to establish an archimedean analogue of the above inequalities, which applies to matrix polynomials with coefficients in the field of complex numbers, equipped with the modulus as its valuation. This allows us in particular to improve the accuracy of the numerical computation of the eigenvalues of such matrix polynomials.

In [15], with Meisam Sharify (IPM, Tehran, Iran), we also establish log-majorization inequalities of the eigenvalues of matrix polynomials using the tropical roots of some scalar polynomials depending only on the norms of the matrix coefficients. This extends to the case of matrix polynomials some bounds obtained by Hadamard, Ostrowski and Pólya for the roots of scalar polynomials.

These works have been presented in [22].

7.4.2. Tropicalization of the central path and application to the complexity of interior point methods

Participants: Xavier Allamigeon, Stéphane Gaubert.

This work is in collaboration with Pascal Benchimol (now with EDF Labs) and Michael Joswig (TU-Berlin).

In optimization, path-following interior point methods are driven to an optimal solution along a trajectory called the central path. The *central path* of a linear program $\mathrm{LP}(A,b,c) \equiv \min\{c\cdot x \mid Ax \leq b,\ x \geq 0\}$ is defined as the set of the optimal solutions (x^μ, w^μ) of the barrier problems:

$$\begin{aligned} \text{minimize} \quad & c\cdot x - \mu(\sum_{j=1}^{n} \log x_j + \sum_{i=1}^{m} \log w_i) \\ \text{subject to} \quad & Ax + w = b,\ x > 0,\ w > 0 \end{aligned}$$

While the complexity of interior point methods is known to be polynomial, an important question is to study the number of iterations which are performed by interior point methods, in particular whether it can be bounded by a polynomial in the dimension (mn) of the problem. This is motivated by one of Smale's problems, on the existence of a strongly polynomial complexity algorithm for linear programming. So far, this question has been essentially addressed though the study of the curvature of the central path, which measures how far a path differs from a straight line, see [75], [74], [77], [76]. In particular, by analogy with the classical Hirsch conjecture, Deza, Terlaky and Zinchencko [76] conjectured that $O(m)$ is also an upper bound for the total curvature.

In a work of X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig, we study the tropicalization of the central path. The *tropical central path* is defined as the logarithmic limit of the central paths of a parametric family of linear programs $\mathrm{LP}(A(t),b(t),c(t))$, where the entries $A_{ij}(t)$, $b_i(t)$ and $c_j(t)$ are definable functions in an o-minimal structure called the *Hardy field*.

A first contribution is to provide a purely geometric characterization of the tropical central path. We have shown that the tropical analytic center is the greatest element of the tropical feasible set. Moreover, any point of the tropical central path is the greatest element of the tropical feasible set intersected with a sublevel set of the tropical objective function.

Thanks to this characterization, we identify a class of path-following interior-point methods which are not strongly polynomial. This class corresponds to primal-dual interior-point methods which iterates in the so-called "wide" neighborhood of the central path arising from the logarithmic barrier. It includes short step, long step as well as predictor-corrector types of interior-point methods. In more details, we establish a lower bound on the number of iterations of these methods, expressed in terms of the number of tropical segments constituting the tropical central path. In this way, we exhibit a family of linear programs with $3d+1$ inequalities in dimension $2d$ on which the aforementioned interior point methods require $\Omega(2^d)$ iterations. The same family provides a counterexample to Deza, Terlaky and Zinchenko's conjecture, having a total curvature in $\Omega(2^d)$.

A first part of these results is in the preprint [61], further results been presented in [32].

7.4.3. Tropical approach to semidefinite programming

Participants: Xavier Allamigeon, Stéphane Gaubert, Mateusz Skomra.

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.

We address 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 first study 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 gathered in the preprint [49].

Then, we show 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. Part of this latter work has been published in the proceedings of the conference ISSAC 2016 [29].

7.5. Applications

7.5.1. Geometry of the Loewner order and application to the synthesis of quadratic invariants in static analysis of program

Participants: Xavier Allamigeon, Stéphane Gaubert, Nikolas Stott.

This work is in collaboration with Éric Goubault and Sylvie Putot (from LIX).

We introduce a new numerical abstract domain based on ellipsoids designed for the formal verification of switched linear systems. The novelty of this domain does not consist in the use of ellipsoids as abstractions, but rather in the fact that we overcome two key difficulties which so far have limited the use of ellipsoids in abstract interpretation. The first issue is that the ordered set of ellipsoids does not constitute a lattice. This implies that there is a priori no canonical choice of the abstraction of the union of two sets, making the analysis less predictable as it relies on the selection of good upper bounds. The second issue is that most recent works using on ellipsoids rely on LMI methods. The latter are efficient on moderate size examples but they are inherently limited by the complexity of interior point algorithms, which, in the case of matrix inequality problems, do not scale as well as for linear programming or second order cone programming problems.

We developed a new approach, in which we reduce the computation of an invariant to the determination of a fixed point, or eigenvector, of a non-linear map that provides a safe upper-approximation of the action induced by the program on the space of quadratic forms. This allows one to obtain invariants of systems of sized inaccessible by LMI methods, at the price of a limited loss of precision. A key ingredient here is the fast computation of least upper bounds in Löwner ordering, by an algebraic algorithm. This relies on the study of the geometry of the space of quadratic forms (Section 7.2.3).

A first part of this work is described in the article [16], which is the extended version of [65] which won the best paper award at the conference EMSOFT 2015. Followup work is dealing with the extension of these results to switched affine systems with guards.

7.5.2. *Performance evaluation of an emergency call center based on tropical polynomial systems*

Participants: Xavier Allamigeon, Vianney Boeuf, Stéphane Gaubert.

This work arose from a question raised by Régis Reboul from Préfecture de Police de Paris (PP), regarding the analysis of the projected evolution of the treatment of emergency calls (17-18-112). This work benefited from the help of LtL Stéphane Raclot, from Brigade de Sapeurs de Pompiers de Paris (BSPP). It is part of the PhD work of Vianney Bœuf, carried out in collaboration with BSPP.

We introduced an algebraic approach which allows to analyze the performance of systems involving priorities and modeled by timed Petri nets. Our results apply to the class of Petri nets in which the places can be partitioned in two categories: the routing in certain places is subject to priority rules, whereas the routing at the other places is free choice.

In [62], we introduced a discrete model, showing that the counter variables, which determine the number of firings of the different transitions as a function of time, are the solutions of a piecewise linear dynamical system. Moreover, we establish that in the fluid approximation of this model, the stationary regimes are precisely the solutions of a set of lexicographic piecewise linear equations, which constitutes a polynomial system over a tropical (min-plus) semifield of germs.

In essence, this result shows that computing stationary regimes reduces to solving tropical polynomial systems. Solving tropical polynomial systems is one of the most basic problems of tropical geometry. The latter provides insights on the nature of solutions, as well as algorithmic tools. In particular, the tropical approach allows one to determine the different congestion phases of the system.

We applied this approach to a case study relative to the project led by Préfecture de Police de Paris, involving BSPP, of a new organization to handle emergency calls to Police (number 17), Firemen (number 18), and untyped emergency calls (number 112), in the Paris area. We initially introduced, in [62], a simplified model of emergency call center, and we concentrated on the analysis of an essential feature of the organization: the two level emergency procedure. Operators at level 1 initially receive the calls, qualify their urgency, handle the non urgent ones, and transfer the urgent cases to specialized level 2 operators who complete the instruction. We solved the associated system of tropical polynomial equations and arrived at an explicit computation of the different congestion phases, depending on the ratio of the numbers of operators of level 2 and 1.

We subsequently developed a more complex model, taking into account the different characteristics of the calls to 17 and 18, and developed a realistic simulation tool to validate the results. Moreover, in [28], we developed an alternative model, relying on fluid Petri nets (dynamical systems with piecewise affine vector fields). We showed that the fluid and discrete models have the same stationary regimes, and that some pathological features of the discrete model (anomalous periodic orbits appearing under certain arithmetical conditions) vanish in the fluid Petri net case.

7.5.3. *Smart Data Pricing*

Participants: Marianne Akian, Jean-Bernard Eytard.

This work is in collaboration with Mustapha Bouhtou (Orange Labs).

The PhD work of Jean-Bernard Eytard concerns the optimal pricing of data trafic in mobile networks. We developed a bilevel programming approach, allowing to an operator to balance the load in the network through price incentives. We showed that a subclass of bilevel programs can be solved in polynomial time, by combining methods of tropical geometry and of discrete convexity. This work has been presented in [31].

AMIB Project-Team

5. New Results

5.1. RNA design

We extended our previous results on RNA design [29], obtained in collaboration with J. Hales, J. Manuch and L. Stacho (Simon Fraser University/Univ. British Columbia, Canada).

Our results provided complete characterizations for the structures that can be designed using restricted alphabets. We provided a complete characterization of designable structures without unpaired bases. When unpaired bases are allowed, we provided partial characterizations for classes of designable/undesignable structures, and showed that the class of designable structures is closed under the stutter operation. Membership of a given structure to any of the classes can be tested in linear time and, for positive instances, a solution could be found in linear time. Finally, we considered a structure-approximating version of the problem that allows to extend helices and, assuming that the input structure avoids two motifs, we provided a linear-time algorithm that produces a designable structure with at most twice more base pairs than the input structure, as illustrated by Fig. 3 .

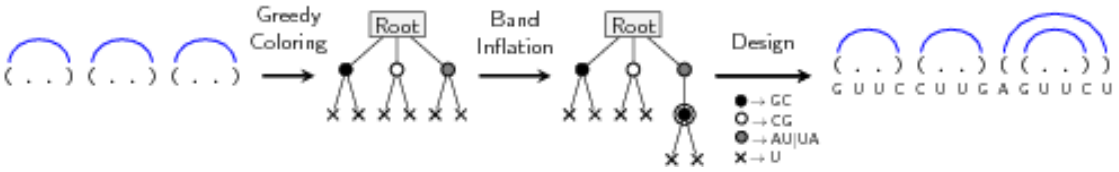


Figure 3. Principle of our structure-approximating version of RNA design: Starting from a potentially undesignable structure, a greedy coloring can be performed and corrected such that the final structure is provably designable in linear time.

In a manuscript accepted for publication in Algorithmica [4], we have shown that our previous results [29] hold for more sophisticated energy models where base-pairs are associated with arbitrary energy contributions. This result, which required a complete overhaul of our previous proofs (e.g. using arguments based on graph coloring), allows us to foresee an extension of (at least some of) our results to state-of-the-art models, such as the Turner energy model.

We also initiated a collaboration with Danny Barash's group at Ben-Gurion university (Israel). We contributed a review of existing tools and techniques for RNA design, to appear as an article within the *Briefings in Bioinformatics* series [2]. We also combined previously contributed methods for design into a new method and web-server for the design of RNAs [3]. This collaboration stemed from the observation that `IncaRNAtion` [36], a random generation algorithm for RNA design recently developed in collaboration with Jérôme Waldispühl's group at McGill University (Montreal, Canada), produced excellent starting points (seed) for classic algorithms based on local-search. In particular, the combination of `IncaRNAtion` and `RNAfbInv` [45] was found to yield particularly promising candidates for design.

5.2. Algorithmics and combinatorics of motifs occurrences

We have developed a new algorithm to compute minimal absent words in external memory. Minimal absent words are used in sequence comparison [23] or to detect biologically significant events. For instance, it was

shown that there exist three minimal words in Ebola virus genomes which are absent from human genome [42]. The identification of such specific-species sequences may prove to be useful for the development of diagnosis and therapeutics. We have already provide an $O(n)$-time and $O(n)$-space algorithm to compute minimal absent words, with an implementation that can be executed in parallel. However these implementations require a large amount of RAM, thus they cannot be used for the human genome on a desktop computer. In our new contribution we developed an external memory implementation, it can compute minimal absent words of length at most 11 for the human genome using only 1GB of RAM in less than 4 hours (manuscript submitted [16]).

Repetitive patterns in genomic sequences have a great biological significance. This is a key issue in several genomic problems as many repetitive structures can be found in genomes. One may cite microsatellites, retrotransposons, DNA transposons, long terminal repeats (LTR), long interspersed nuclear elements (LINE), ribosomal DNA, short interspersed nuclear elements (SINE). Knowledge about the length of a maximal repeat also has algorithmic implications, most notably the design of assembly algorithms that rely upon de Bruijn graphs.

Analytic combinatorics allowed us to derive formula for the expected length of repetitions in a random sequence [9]. The originality of the approach is the demonstration of a Large Deviation principle and the use of Lagrange multipliers. This allowed for a generalization of previous works on a binary alphabet. Simulations on random sequences confirmed the accuracy of our results. As an application, the sample case of Archaea genomes illustrated how biological sequences may differ from random sequences, and in turns provides tools to extract repetitive sequences.

5.3. Integrative RNA structural modeling

To circumvent expensive, low-throughput, 3D experimental techniques such as X-ray crystallography, a low resolution/high throughput technology called SHAPE is increasingly favored for structural modeling by structural biologists.

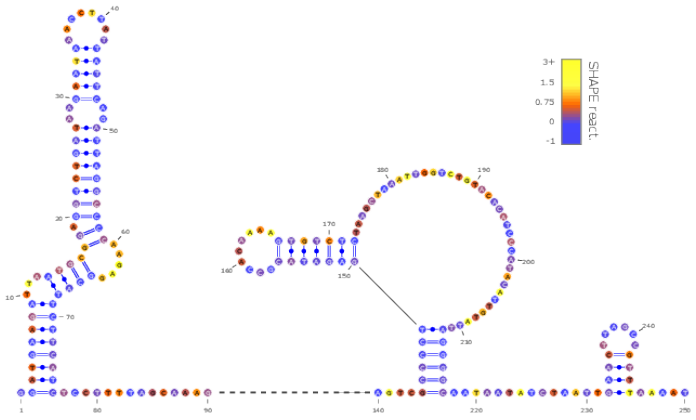
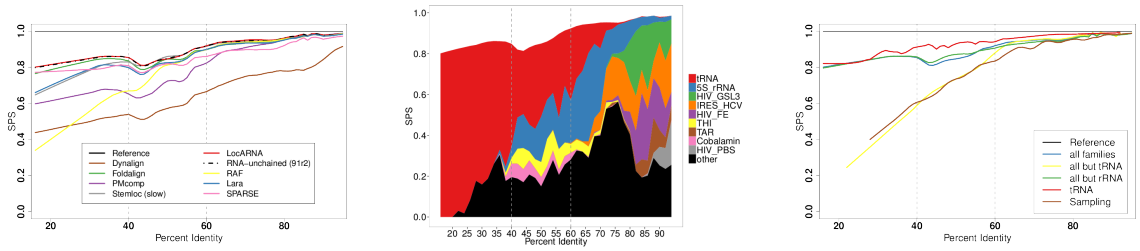


Figure 4. Conserved and thermodynamically-stable structure elements revealed by our analysis of an Ebola UTR region.

Within Afaf Saaidi's thesis, funded by the *Fondation pour la Recherche Médicale* and co-supervised by Bruno Sargueil at Faculté de Pharmacie of Université Paris V, we have developed integrative modeling strategies based on Boltzmann sampling. Preliminary results, obtained by applying these methods to model the structures of 3'UTR regions in Ebola, were presented at JOBIM 2016 [14].

Moreover, in collaboration with McGill University (Canada), we cross-examined mutate-and-map data (MaM [30]) in the light of evolutionary data. MaM data consist in the sequential SHAPE probing of a set of mutant RNAs, obtained by systematic point-wise mutations, to highlight structurally-dependent nucleotides, later to use dependent pairs as constraints in (an automated) structural modeling. We chose to adopt an alternative perspective on MaM data, and used the perturbation of the SHAPE profiles as a proxy for the structural disruption induced by a mutation. Disruptive mutations are rescued within homologs, *i.e.* compensated to re-establish the structure. However, our analysis also revealed the existence of non-structurally local (neither on the 2D or 3D levels) nucleotides which have significant mutual-information with highly disruptive positions, despite not being involved in any obvious compensatory relationship.

We hypothesized that such mutations are revealing of interactions involving RNA. In a manuscript published in *Nucleic Acids* journal, we tested and validated this hypothesis by showing its capacity to discriminate discriminative positions that are known to be in contact with specific ligands (proteins, DNA, small molecules...) [10].



a. *b.* *c.*

Figure 5. Typical software for comparative RNA structure prediction exhibit a dent in performance within the 40%-60% sequence identity range when benchmarked using the popular Bralibase data set (a.). However, this is due to the overrepresentation of a well-predicted type of RNA (tRNAs, red area) for low-identity ranges (b.). A re-evaluation of state-of-the-art software on an unbiased (c., brown line) reveals much more modest predictive capacities than initially believed in the community.

A fruitful line of research for RNA structure prediction is based on a comparative approach. Whenever homologous RNAs are identified, a classic strategy is to perform a simultaneous alignment and folding of several RNAs. Many software (30+) have been contributed over the past decades for this problem, leading to the introduction of benchmarks, one of the most prominent being the `Bralibase`, to position new developments and identify axes of progression. One such desired improvement, as illustrated in Figure 5 , was the difficulties experienced by most software around the 40-60% sequence identity range, which was believed to arise from deep algorithmic reasons. In collaboration with Cedric Chauve (Simon Fraser University, Canada) Benedikt Löwes and Robert Giegerich (Bielefeld University, Germany), we showed that this perceived difficulty was simply the consequence of a strong bias towards tRNAs in the 40-60% sequence identity region. Moreover, we argued that the overall performance of existing tools for low sequence identities were largely overestimated [8].

Finally, we presented at JOBIM 2016 an efficient implementation, called `LicoRNA`, of our parameterized complexity algorithm based on tree-decomposition for the sequence/structure alignment of RNA [15]. Specifically, we showed that our `LicoRNA`, by including an expressive scoring scheme and capturing pseudoknots of arbitrary complexity, generally outperforms previously contributions for the problem.

5.4. Combinatorial foundations

Pairwise ordered tree alignment are combinatorial objects that appear in RNA secondary structure comparison. However, the usual representation of tree alignments as supertrees is ambiguous, i.e. two distinct supertrees

may induce identical sets of matches between identical pairs of trees. This ambiguity is uninformative, and detrimental to any probabilistic analysis. In a recent collaboration with Cédric Chauve (SFU Vancouver, Canada) and Julien Courtiel (LIPN, Paris XII) presented at the ALCOB'16 conference, we considered tree alignments up to equivalence [11]. Our first result was a precise asymptotic enumeration of tree alignments, obtained from a context-free grammar by means of basic analytic combinatorics. Our second result focused on alignments between two given ordered trees. By refining our grammar to align specific trees, we obtained a decomposition scheme for the space of alignments, and used it to design an efficient dynamic programming algorithm for sampling alignments under the Gibbs-Boltzmann probability distribution. This generalizes existing tree alignment algorithms, and opens the door for a probabilistic analysis of the space of suboptimal RNA secondary structures alignments.

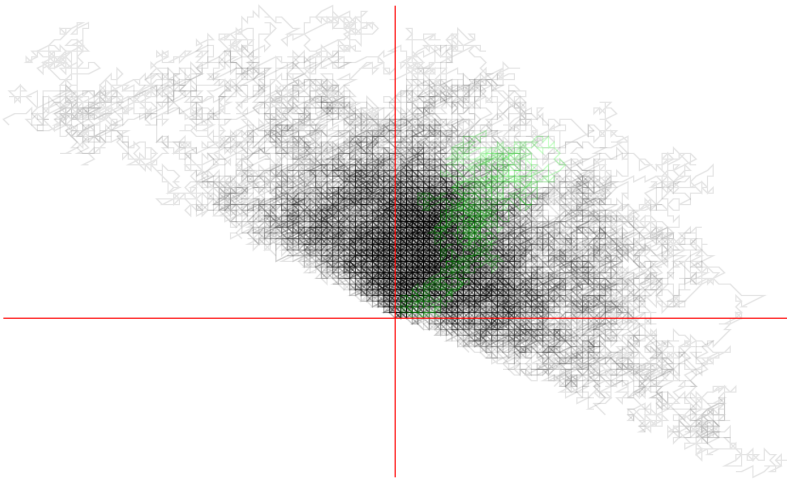


Figure 6. Random 2D walks (green walks) confined in the positive can be generated efficiently by performing rejection from a well-chosen 1D model (black walks) [12].

Finally, in collaboration with Marni Mishna (Simon Fraser University, Canada) and Jérémie Lumbroso (Princeton University, USA), we considered the uniform random generation of *difficult*, or *reluctant*, 2D discrete walks that remain confined in the positive quarter plane. We proposed a naive dynamic programming algorithm having complexity $O(n^4)$ for any step set. We also exploited the remark that any quarterplane walks can be transformed into a well-chosen 1D model having the same exponential growth factor. However, such a 1D model takes irrational steps, leading us to explore new avenues for the random generation. This work was presented at the GASCOM'16 conference [12].

5.5. Comparative genomics

D. Iakovishina defended in 2015 a PhD thesis co-advised by M. Régnier and V. Boeva (Curie Institute). She proposed a new computational method to detect structural variants using whole genome sequencing data. It combines two techniques that are based either on the detection of paired-end mapping abnormalities or on the detection of the depth of coverage. SV-BAY relies on a probabilistic Bayesian approach and includes a modelization of possible sequencing errors, read mappability profile along the genome and changes in the GC-content. Keeping only somatic SVs is an additional option when matched normal control data are provided. SV-BAY compares favorably with existing tools on simulated and experimental data sets [6] Software SV-BAY is freely available https://github.com/InstitutCurie/SV-Bay.

As a side product, a novel exhaustive catalogue of SV types -to date the most comprehensive SV classification- was built. On the grounds of previous publications and experimental data, seven new SV types, ignored by the existing SV calling algorithms, were exhibited.

We also contributed, in collaboration with Céline Scornavacca's group (ISEM, Montpellier) to the algorithmic foundations of the `EcceTERA` software [7] for the reconciliation of gene and species phylogenetic trees. This software adopts a maximum parsimony approach to predict in an evolutionary model that includes duplications, losses and transfers of genes.

GALEN Project-Team

7. New Results

7.1. Learning Grammars for Architecture-Specific Facade Parsing

Paticipants: Nikos Paragios (in collaboration with researchers from Université Paris-Est, LIGM, ENPC)

In [5], we present a novel framework to learn a compact grammar from a set of ground-truth images. To this end, parse trees of ground-truth annotated images are obtained running existing inference algorithms with a simple, very general grammar. From these parse trees, repeated subtrees are sought and merged together to share derivations and produce a grammar with fewer rules. Furthermore, unsupervised clustering is performed on these rules, so that, rules corresponding to the same complex pattern are grouped together leading to a rich compact grammar.

7.2. Non-Rigid Surface Registration

Paticipants: Dimitris Samaras, Nikos Paragios

This work [13] casts surface registration as the problem of finding a set of discrete correspondences through the minimization of an energy function, which is composed of geometric and appearance matching costs, as well as higher-order deformation priors. Two higher-order graph-based formulations are proposed under different deformation assumptions.

7.3. Monocular Surface Reconstruction using 3D Deformable Part Models

Paticipants: Maxim Berman, Stefan Kinauer, Iasonas Kokkinos

In this work [22] we train and detect part-based object models in 2D images, recovering 3D position and shape information (per part positions), allowing for a 3D reconstruction of the object. The resulting optimization problem is solved via a Branch&Bound approach, yielding detection results within a fraction of a second.

7.4. Learning with Non-modular loss functions

Paticipants: Jiaqian Yu, Matthew Blaschko

We have proposed an alternating direction method of multipliers (ADMM) based decomposition method loss augmented inference, that only depends on two individual solvers for the loss function term and for the inference term as two independent subproblems. In this way, we can gain computational efficiency and achieve more flexibility in choosing our non-modular loss functions of interest. We have proposed a novel supermodular loss function that empirically achieved better performance on the boundary of the objects, finding elongated structure [33]. We also introduced a novel convex surrogate operator for general non-modular loss functions, which provides for the first time a tractable solution for loss functions that are neither supermodular nor submodular, e.g. Dice loss. This surrogate is based on a canonical submodular-supermodular decomposition for which we have demonstrated its existence and uniqueness. It is further proven that this surrogate is convex, piecewise linear, an extension of the loss function, and for which subgradient computation is polynomial time [32][31].

7.5. Asymptotic Variance of MMD and Relative MMD

Paticipants: Eugene Belilovsky, Wacha Bounliphone, Matthew Blaschko (in collaboration with researchers at UCL and Deepmind)

Kernel mean embeddings allow for comparisons of complex distributions. They have been recently heavily used in hypothesis testing to compare distributions as well as in the nascent field of deep generative modeling. In this work we derived the asymptotic variance of the MMD and the cross covariance between joint MMD. We showed how this can be used effectively for model selection in complex Deep Generative Models where the likelihood metric is not accessible. Our results on the asymptotic variance of the MMD have already been used by other researchers to propose an efficient method for optimal testing and improved training of generative models.

7.6. Deconvolution and Deinterlacing of Video Sequences

Paticipants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with F. Abboud, PhD student, J.-H. Chenot and L. Laborelli, research engineers, Institut National de l'Audiovisuel)

Optimization methods play a central role in the solution of a wide array of problems encountered in various application fields, such as signal and image processing. Especially when the problems are highly dimensional, proximal methods have shown their efficiency through their capability to deal with composite, possibly non smooth objective functions. The cornerstone of these approaches is the proximity operator, which has become a quite popular tool in optimization. In this work, we propose new dual forward-backward formulations for computing the proximity operator of a sum of convex functions involving linear operators. The proposed algorithms are accelerated thanks to the introduction of a block coordinate strategy combined with a preconditioning technique. Numerical simulations emphasize the good performance of our approach for the problem of jointly deconvoluting and deinterlacing video sequences.

7.7. A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise

Paticipants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Y. Marnissi, PhD student at Univ. Paris-Est Marne la Vallée and Y. Zheng, IBM Research China)

In this work, a methodology is investigated for signal recovery in the presence of non-Gaussian noise. In contrast with regularized minimization approaches often adopted in the literature, in our algorithm the regularization parameter is reliably estimated from the observations. As the posterior density of the unknown parameters is analytically intractable, the estimation problem is derived in a variational Bayesian framework where the goal is to provide a good approximation to the posterior distribution in order to compute posterior mean estimates. Moreover, a majorization technique is employed to circumvent the difficulties raised by the intricate forms of the non-Gaussian likelihood and of the prior density. We demonstrate the potential of the proposed approach through comparisons with state-of-the-art techniques that are specifically tailored to signal recovery in the presence of mixed Poisson-Gaussian noise. Results show that the proposed approach is efficient and achieves performance comparable with other methods where the regularization parameter is manually tuned from an available ground truth.

7.8. The Majorize-Minimize Subspace Algorithm and Block Parallelization

Paticipants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with S. Cadoni, Master student at Univ. Paris-Est Marne la Vallée and Dr C. Chaux, Univ. Aix-Marseille)

State-of-the-art methods for solving smooth optimization problems are nonlinear conjugate gradient, low memory BFGS, and Majorize-Minimize (MM) subspace algorithms. The MM subspace algorithm which has been introduced more recently has shown good practical performance when compared with other methods on various optimization problems arising in signal and image processing. However, to the best of our knowledge, no general result exists concerning the theoretical convergence rate of the MM subspace algorithm. The paper [3] aims at deriving such convergence rates both for batch and online versions of the algorithm and, in particular, discusses the influence of the choice of the subspace. We also propose a Block Parallel Majorize-Minimize Memory Gradient (BP3MG) algorithm for solving large scale optimization problems in [16]. This

algorithm combines a block coordinate strategy with an efficient parallel update. The proposed method is applied to a 3D microscopy image restora- tion problem involving a depth-variant blur, where it is shown to lead to significant computational time savings with respect to a sequential approach.

7.9. Stochastic Forward-Backward and Primal-Dual Approximation Algorithms with Application to Online Image Restoration

Paticipants: Jean-Christophe Pesquet (In collaboration with Pr. P. L. Combettes, North Carolina State university)

Stochastic approximation techniques have been used in various contexts in data science. We propose a stochastic version of the forward-backward algorithm for minimizing the sum of two convex functions, one of which is not necessarily smooth. Our framework can handle stochastic approximations of the gradient of the smooth function and allows for stochastic errors in the evaluation of the proximity operator of the nonsmooth function. The almost sure convergence of the iterates generated by the algorithm to a minimizer is established under relatively mild assumptions. We also propose a stochastic version of a popular primal-dual proximal splitting algorithm, establish its convergence, and apply it to an online image restoration problem.

7.10. Random primal-dual proximal iterations for sparse multiclass SVM

Paticipants: Jean-Christophe Pesquet (in collaboration with Pr. G. Chierchia, Univ. Paris-Est Marne la Vallée, and Dr. N. Pustelnik, ENS Lyon)

Sparsity-inducing penalties are useful tools in variational methods for machine learning. In this paper, we propose two block-coordinate descent strategies for learning a sparse multiclass support vector machine. The first one works by selecting a subset of features to be updated at each iteration, while the second one performs the selection among the training samples. These algorithms can be efficiently implemented thanks to the flexibility offered by recent randomized primal-dual proximal methods. Experiments carried out for the supervised classification of handwritten digits demonstrate the interest of considering the primal-dual approach in the context of block-coordinate descent. The efficiency of the proposed algorithms is assessed through a comparison of execution times and classification errors.

7.11. PALMA, an improved algorithm for DOSY signal processing

Paticipants: Emilie Chouzenoux (in collaboration with Prof. M.-A. Delsuc, IGBMC, Strasbourg, and A. Cherni, PhD student, Univ. Strasbourg)

NMR is a tool of choice for the measure of diffusion coefficients of species in solution. The DOSY experiment, a 2D implementation of this measure, has proven to be particularly useful for the study of complex mixtures, molecular interactions, polymers, etc. However, DOSY data analysis requires to resort to inverse Laplace transform, in particular for polydisperse samples. This is a known difficult numerical task, for which we present here a novel approach. A new algorithm based on a splitting scheme and on the use of proximity operators is introduced. Used in conjunction with a Maximum Entropy and λ_1 hybrid regularisation [39], this algorithm converges rapidly and produces results robust against experimental noise. This method has been called PALMA. It is able to reproduce faithfully monodisperse as well as polydisperse systems, and numerous simulated and experimental examples are presented in [35]. It has been implemented on the server [http://palma.labo.igbmc.fr] where users can have their datasets processed automatically.

7.12. Graph-based change detection and classification in satellite image pairs

Paticipants: Maria Vakalopoulou, Nikos Paragios

We proposed a scalable, modular, metric-free, single-shot change detection/registration method for remote sensing image pairs [11]. The framework exploits a decomposed interconnected graphical model formulation where in the presence of changes the iconic similarity constraints are relaxed. We employ a discretized, grid-based deformation space. State-of-the-art linear programming and duality principles have been used to optimize the joint solution space where local consistency is imposed on the deformation and the detection space. The proposed framework is working both in a unsupervised and supervised manner depending on the application. The developed method has been validated through large scale experiments on several multi-temporal very high resolution optical satellite datasets. Also a novel generic framework has been designed, developed and validated for addressing simultaneously the tasks of image registration, segmentation and change detection from multisensor, multiresolution, multitemporal satellite image pairs [30]. Our approach models the inter-dependencies of variables through a higher order graph. A patch-based deep learning strategy has been employed and used for segmentation likelihoods. The evaluation of the developed framework was performed on the '2016 IEEE GRSS Data Fusion Contest' dataset and indicate very promising results for all three different tasks.

7.13. Graphical models in artificial vision

Paticipants: Nikos Komodakis, M. Pawan Kumar, Stavros Alchatzidis, Enzo Ferrante, Evangelia Zacharaki, Nikos Paragios

Computer vision tasks are often reformulated as mathematical inference problems where the objective is to determine the set of parameters corresponding to the lowest potential of a task-specific objective function. Graphical models have been the most popular formulation in the field over the past two decades. In [7] we focus on the inference component of the problem and in particular we discuss in a systematic manner the most commonly used optimization principles in the context of graphical models. In [8] we briefly review hyper-graph representations as prominent tools in the casting of perception as a graph optimization problem. We discuss their strength and limitations, provide appropriate strategies for their inference and present their application to address a variety of problems in biomedical image analysis.

Multi-atlas segmentation has emerged in recent years as a simple yet powerful approach in medical image segmentation. It commonly comprises two steps: (1) a series of pairwise registrations that establish correspondences between a query image and a number of atlases, and (2) the fusion of the available segmentation hypotheses towards labeling objects of interest. In [2], we introduce a novel approach that solves simultaneously for the underlying segmentation labels and the multi-atlas registration. We propose a pairwise Markov Random Field approach,where registration and segmentation nodes are coupled towards simultaneously recovering all atlas deformations and labeling the query image.

7.14. Pattern analysis of EEG signals with epileptic activity

Paticipants: Evangelia Zacharaki (in collaboration with Prof. M. Megalooikonomou, University of Patras and M. Koutroumanidis, King's College, London)

We have addressed the needs of epileptic patients and healthcare professionals, aiming at the design and development of a non-intrusive personal health system for the monitoring and analysis of epilepsy-relevant multi-parametric data and the documentation of the epilepsy related symptoms. Specifically, we investigated the classification of epileptic and non-epileptic events from EEG based on temporal and spectral analysis and different fusion schemes [9]. We also studied the EEG brain activity during whole night sleep, since sleep is recognized as a major precipitator of epileptic activity [12].

LIFEWARE Project-Team

7. New Results

7.1. Analog computation in the cell: specifications, compilation into biochemical reactions and computational complexity

Participants: François Fages, Guillaume Le Guludec.

The continuous nature of many protein interactions leads us to consider mixed analog-digital computation methods, for which 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 [18], we derive from these results a Turing completeness result for elementary reaction systems (without polymerization) under the differential semantics, and present a compiler of behavioural specifications into biochemical reactions which can be compared to natural circuits acquired through evolution. We illustrate this approach through the example of the MAPK signaling module which has a function of analog-digital converter in the cell, and through the cell cycle control.

The biochemical compiler is implemented in BIOCHAM v4.0 which will be soon released. We plan to use it in the ANR-MOST project BIOPSY on "Biochemical Programming" for the design of artificial biosensors and the programming of (non-living) protocells in collaboration with Franck Molina, CNRS Sys2diag lab, Montpellier.

7.2. Influence systems vs reaction systems

Participants: François Fages, Thierry Martinez, David Rosenblueth, Sylvain Soliman, Denis Thieffry.

In Systems Biology, modelers develop more and more reaction-based models to describe the mechanistic biochemical reactions underlying cell processes. They may also work, however, with a simpler formalism of influence graphs, to merely describe the positive and negative influences between molecular species. The first approach is promoted by reaction model exchange formats such as SBML, and tools like CellDesigner, while the second is supported by other tools that have been historically developed to reason about boolean gene regulatory networks. In practice, modelers often reason with both kinds of formalisms, and may find an influence model useful in the process of building a reaction model. In [11], we introduce a formalism of influence systems with forces, and put it in parallel with reaction systems with kinetics, in order to develop a similar hierarchy of boolean, discrete, stochastic and differential semantics. We show that the expressive power of influence systems is the same as that of reaction systems under the differential semantics, but weaker under the other interpretations, in the sense that some discrete behaviours of reaction systems cannot be expressed by influence systems. This approach leads us to consider a positive boolean semantics which we compare to the asynchronous semantics of gene regulatory networks à la Thomas. We study the monotonicity properties of the positive boolean semantics and derive from them an efficient algorithm to compute attractors.

7.3. What population reveals about individual cell identity: Single-cell parameter estimation of models of gene expression in yeast

Participants: Grégory Batt, Pascal Hersen, Artémis Llamosi.

[0] Amaury Pouly, "Continuous models of computation: from computability to complexity", PhD Thesis, Ecole Polytechnique, Nov. 2015.

Significant cell-to-cell heterogeneity is ubiquitously observed in isogenic cell populations. Consequently, parameters of models of intracellular processes, usually fitted to population-averaged data, should rather be fitted to individual cells to obtain a population of models of similar but non-identical individuals. In [6], we propose a quantitative modeling framework that attributes specific parameter values to single cells for a standard model of gene expression. We combine high quality single-cell measurements of the response of yeast cells to repeated hyperosmotic shocks and state-of-the-art statistical inference approaches for mixed-effects models to infer multidimensional parameter distributions describing the population, and then derive specific parameters for individual cells. The analysis of single-cell parameters shows that single-cell identity (e.g. gene expression dynamics, cell size, growth rate, mother-daughter relationships) is, at least partially, captured by the parameter values of gene expression models (e.g. rates of transcription, translation and degradation). Our approach shows how to use the rich information contained into longitudinal single-cell data to infer parameters that can faithfully represent single-cell identity. This is the first demonstration that biologically-meaningful values for parameter of intracellular processes can be attributed to individual cells.

7.4. The cost of cellular adaptation to stress: tradeoff between short-term and long-term adaptation to osmotic stress in yeast

Participants: Grégory Batt, Ewen Corre, Pascal Hersen, Artémis Llamosi.

Upon stress, cells have evolved complex adaptation strategies to environmental variations which include sensing, information processing and modification of metabolic and transcriptional activity. The reaction of yeast cells to hyperosmotic stress spans several timescales and includes massive gene-expression changes, bio-compatible osmolyte production, and direct action on the cell cycle. Despite a detailed knowledge of molecular events, the impact of stress-response on cellular resources is poorly known. In particular, strong and fast adaptation which alter proliferation in the short term while conferring advantage on the long term are important drivers of stress-response evolution.

In this study, we use microfluidics to vary dynamically both the source of cost (osmotic stress) and the available metabolic resources (glucose) while monitoring cellular proliferation. We show that, under hyper-osmotic stress, metabolic resources are rerouted towards the production of glycerol through activation of an essential enzyme in glycerol production. This reveals the nature of the burden imposed by osmotic stress and, more generally, allows us to better understand the evolutionary tradeoffs between stress response and proliferation.

7.5. Balancing a genetic inverted pendulum

Participants: Grégory Batt, Pascal Hersen, Jean-Baptiste Lugagne, Jean-Baptiste Caron.

The ability to routinely control complex genetic circuits in vivo and in real-time promises quantitative understanding of cellular processes of unprecedented precision and quality. With combined efforts in microfluidic design, microscope automation, image analysis, modeling and control theory, we propose a platform for real-time, single-cell, *in silico* control of genetic networks in *E. coli*. The circuit we are trying to control is a genetic toggle switch, a foundational circuit in synthetic biology, which consists of two genes that repress each other. This genetic system features two stable equilibrium points where one of the genes has taken over. Our objective is to dynamically balance the circuit in single cells around a third, unstable equilibrium point at which no gene dominates and their mutual repression strengths are balanced. This is similar to the landmark problem in control theory of stabilizing an inverted pendulum in its upright position. Although our work indicates that this real-time control approach can drive convoluted genetic networks towards states that are inaccessible to traditional genetic perturbations such as knock-outs and promoter induction, the a priori quantitative knowledge of the system required for achieving this control is minimal. We show that even a simple Proportional-Integral controller can maintain in a balanced state the toggle switch in single cells. Finally, we demonstrate that similar results can be obtained by applying periodic inductions, identical to all cells in the population. Given the fact that all cells behave differently, this result was highly unexpected. It can however be understood as an example of dynamic stabilization, analogous to the solution proposed by Kapitza for the inverted pendulum.

These results are presented in the PhD thesis of Jean-Baptiste Lugagne [2].

7.6. A look-ahead simulation algorithm for DBN models of biochemical pathways

Participants: Grégory Batt, Sucheendra Palaniappan.

Dynamic Bayesian Networks (DBNs) have been proposed as an efficient abstraction formalism of biochemical models. They have been shown to approximate well the dynamics of biochemical models, while offering improved efficiency for their analysis. In [14], we compare different representations and simulation schemes on these DBNs, testing their efficiency and accuracy as abstractions of biological pathways. When generating these DBNs, many configurations are never explored by the underlying dynamics of the biological systems. This can be used to obtain sparse representations to store and analyze DBNs in a compact way. On the other hand, when simulating these DBNs, singular configurations may be encountered, that is configurations from where no transition probability is defined. This makes simulation more complex. We initially evaluate two simple strategies for dealing with singularities: first, re-sampling simulations visiting singular configurations; second filling up uniformly these singular transition probabilities. We show that both these approaches are error prone. Next, we propose a new algorithm which samples only those configurations that avoid singularities by using a look-ahead strategy. Experiments show that this approach is the most accurate while having a reasonable run time.

7.7. Logical model specification aided by model-checking techniques: application to the mammalian cell cycle regulation

Participants: François Fages, Sylvain Soliman, Denis Thieffry, Pauline Traynard.

Understanding the temporal behaviour of biological regulatory networks requires the integration of molecular information into a dynamical model. However, the analysis of model dynamics faces a combinatorial explosion as the number of regulatory components and interactions increases. In [8], we use model-checking techniques to verify sophisticated dynamical properties resulting from the model influence structure in the absence of kinetic assumption. We demonstrate the power of this approach by analysing a Boolean influence model of the molecular network controlling mammalian cell cycle. This approach enables a systematic analysis of model properties, the delineation of model limitations, and the assessment of various refinements and extensions based on recent experimental observations. The resulting logical model accounts for the main irreversible transitions between cell cycle phases, the sequential activation of cyclins, and the inhibitory role of Skp2, and further emphasizes the multifunctional role for the cell cycle inhibitor Rb.

7.8. Model-based investigation of the circadian clock and cell cycle coupling in mouse embryonic fibroblasts: prediction of RevErb–α up-regulation during mitosis

Participants: François Fages, Sylvain Soliman, Pauline Traynard.

Experimental observations have put in evidence autonomous self-sustained circadian oscillators in most mammalian cells, and proved the existence of molecular links between the circadian clock and the cell cycle. Some mathematical models have also been built to assess conditions of control of the cell cycle by the circadian clock. However, recent studies in individual NIH3T3 fibroblasts have shown an unexpected acceleration of the circadian clock together with the cell cycle when the culture medium is enriched with growth factors, and the absence of such acceleration in confluent cells. In [9], in order to explain these observations, we study a possible entrainment of the circadian clock by the cell cycle through a regulation of clock genes around the mitosis phase. We develop a computational model and a formal specification of the observed behavior to investigate the conditions of entrainment in period and phase. We show that either the selective activation of RevErb-α or the selective inhibition of Bmal1 transcription during the mitosis phase, allow us to fit the experimental data on both period and phase, while a uniform inhibition of transcription during mitosis seems incompatible with the phase data. We conclude on the arguments favouring the RevErb-α up-regulation hypothesis and on some further predictions of the model.

7.9. Stochastic continuous optimization backend for the constraint modelling language MiniZinc with applications to geometrical placement problems

Participants: François Fages, Thierry Martinez, Sylvain Soliman.

MiniZinc is a solver-independent constraint modeling language which is increasingly used in the constraint programming community. It can be used to compare different solvers which are currently based on either Constraint Programming, Boolean satisfiability, Mixed Integer Linear Programming, and recently Local Search. In [12], [13] we present a stochastic continuous optimization backend for MiniZinc models over real numbers. More specifically, we describe the translation of FlatZinc models into objective functions over the reals, and their use as fitness functions for the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) solver. We illustrate this approach with the declarative modeling and solving of hard geometrical placement problems [16], motivated by packing applications in logistics [10] involving mixed square-curved shapes and complex shapes defined by Bézier curves.

Beyond these applications to packing problem, our real motivation for these developments is the solving of parameter search problems in computational systems biology and its implementation in BIOCHAM.

7.10. Mixture model CMA-ES

Participants: François Fages, Nicolas Vasselin.

In [19], we report on our attempt to improve the CMA-ES global optimization algorithm based on two ideas: first the use of Sobol's quasi-random low discrepancy numbers instead of pseudo-random numbers, second the design of a mixture model extension of CMA-ES (MM-CMA-ES) which, instead of doing restarts with an important loss of information at each restart, evolves a dynamic set of multivariate normal distributions in parallel, using an EM clustering algorithm at each step to decide of population splittings and mergings. On the standard Coco benchmark for evaluating global stochastic optimization methods, the use of Sobol numbers shows a quite uniform improvement by 30% as was already shown by Teytaud last year [0]. On the other hand, MM-CMA-ES does not show speed-up w.r.t. CMA-ES with IBOP restart strategy, even on objective functions with many local minima such as the Rastragin function. The reason is the overhead in the number of evaluation of the objective functions, introduced by the MM strategy, and the very subtle effect of the adaptive step size strategy of CMA-ES to escape from the covering of several local minima by one (large) normal distribution.

7.11. Metro energy optimization through rescheduling

Participants: François Fages, Thierry Martinez.

The use of regenerative braking is a key factor to reduce the energy consumption of a metro line. In the case where no device can store the energy produced during braking, only the metros that are accelerating at the same time can benefit from it. Maximizing the power transfers between accelerating and braking metros thus provides a simple strategy to benefit from regenerative energy without any other hardware device. In [15], we use a mathematical timetable model to classify various metro energy optimization problems studied in the literature and prove their NP-hardness by polynomial reductions of SAT. We then focus on the problem of minimizing the global energy consumption of a metro timetable by modifying the dwell times in stations. We present a greedy heuristic algorithm which aims at locally synchronizing braking trains along the timetable with accelerating trains in their time neighbourhood, using a non-linear approximation of energy transfers. On a benchmark of the litterature composed of six small size timetables, we show that our greedy heuristics performs better than CPLEX using a MILP formulation of the problem with a linear approximation of the objective function. We also show that it runs ten times faster than a state-of-the-art evolutionary algorithm, called the covariance matrix adaptation evolution strategy (CMA-ES), using the same non-linear objective function on these small size instances. On real data leading to 10000 decision variables on which both MILP and CMA-ES do not provide solutions, our dedicated algorithm computes solutions with a reduction of energy consumption ranging from 5% to 9%.

[0] O. Teytaud. Quasi-random numbers improve the CMA-ES on the BBOB testbed. Artificial Evolution (EA2015), 2015, Lyon, France. Springer Verlag, pp.13

This work done in 2014 in the Cifre PhD Thesis of David Fournier with General Electric Transportation has received this year the Gold Medal of the *Annual Alstom Contest* “I Nove You” in the “Green Innovation” category.

M3DISIM Project-Team

7. New Results

7.1. Mathematical and Mechanical modeling

7.1.1. A 3D contact-mechanics model of the heart and thorax for seismocardiography

Participants: Alexandre Laurin [correspondant], Sébastien Imperiale, Dominique Chapelle, Philippe Moireau.

The current interpretation of seismocardiogram fiducial points depends on their phenomenological association with the timing of events on simultaneous echocardiograms. Signal processing methods can be devised to acquire these timings automatically (see [21] and [22]). So far, no causal framework has been tested to explain this timing, nor their direction and amplitude. This work attempted to adapt a comprehensive 3D cardiac model to interact through contact with a model of the thoracic cage. The heart model was designed to represent multi-scale, multi-physics physiological processes such as the electrical activation, the mechanical contraction, as well as the system circulation. The objective is to link observed acceleration of the sternum to the underlying physiology, and offer a potential mechanical explanation for them. The modelling chain necessary to go from the heart model to a simulated SCG has been successfully implemented (see Figure 1). Furthermore, the complexity of the thoracic model has been substantially reduced, without deteriorating results, to improve the portability of the entire process. Once the relevant parameters of in-vivo thoraces will have been precisely identified, it will be possible to compute heart forces and the various cardiac events that cause them directly from SCG measurements. The subsequent aim is to apply the model to ageing and pathological physiologies.

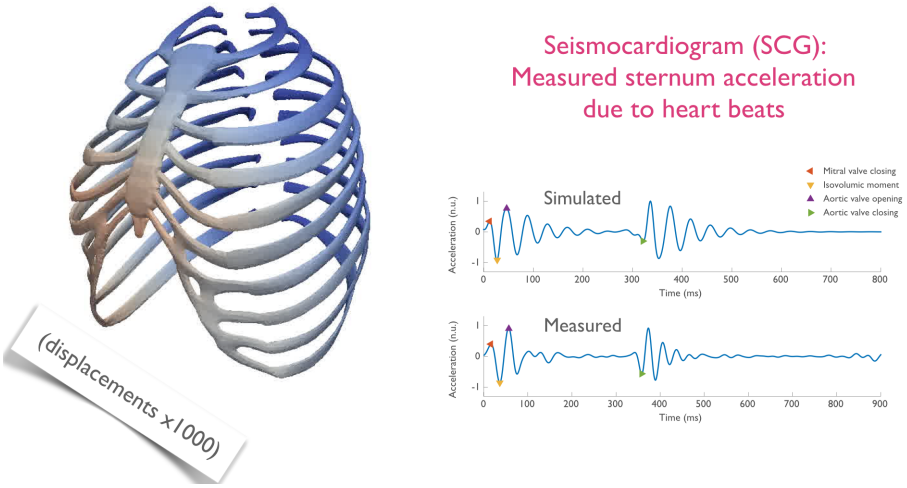


Figure 1. Simulation of the thorax deformation due to a heart beat and of the corresponding seismocardiogram

7.1.2. Multi-scale modeling of muscle contraction

Participants: François Kimmig [correspondant], Dominique Chapelle, Matthieu Caruel.

This works aims at proposing a multi-scale model of the muscular contraction that can be used in the context of cardiac simulation. The modeling will be based on the stochastic equations that describe muscular contraction at the molecular level. Asymptotic counterparts of the stochastic model will be considered in order to provide pertinent simplified models. The modeling elements will be confronted with experiments that will be performed on cardiac muscle cells by collaborators in the team of Professor Vincenzo Lombardi at the University of Florence.

In the framework of this collaboration, a chemicomechanical model is being implemented into CardiacLab, a simulation environment developed by the team. It will enrich the range of modeling tools of the team for the active contribution of muscle cells to the cardiac behavior.

7.1.3. Multiphysics and multiscale modelling, data-model fusion and integration of organ physiology in the clinic: ventricular cardiac mechanics

Participants: Radomir Chabiniok [correspondant], Philippe Moireau, Dominique Chapelle, Maxime Sermesant [Team Asclepios].

With heart and cardiovascular diseases continually challenging healthcare systems worldwide, translating basic research on cardiac (patho)physiology into clinical care is essential. Exacerbating this already extensive challenge is the complexity of the heart, relying on its hierarchical structure and function to maintain cardiovascular flow. Computational modelling has been proposed and actively pursued as a tool for accelerating research and translation. Allowing exploration of the relationships between physics, multiscale mechanisms and function, computational modelling provides a platform for improving our understanding of the heart. Further integration of experimental and clinical data through data assimilation and parameter estimation techniques is bringing computational models closer to use in routine clinical practice. This work published in [17] reviews developments in computational cardiac modelling and how their integration with medical imaging data is providing new pathways for translational cardiac modelling.

7.1.4. Eidolon: visualization and computational framework for multi-modal biomedical data analysis

Participant: Radomir Chabiniok [correspondant].

Biomedical research, combining multi-modal image and geometry data, presents unique challenges for data visualization, processing, and quantitative analysis. Medical imaging provides rich information, from anatomical to deformation, but extracting this to a coherent picture across image modalities with preserved quality is not trivial. Addressing these challenges and integrating visualization with image and quantitative analysis results in Eidolon, a platform which can adapt to rapidly changing research workflows. In the paper [26] we outline Eidolon, a software environment aimed at addressing these challenges, and discuss the novel integration of visualization and analysis components. These capabilities are demonstrated through the example of cardiac strain analysis, showing the Eidolon supports and enhances the workflow.

7.1.5. Mathematical and numerical modeling of elastic waves propagation in the heart

Participants: Federica Caforio [correspondant], Dominique Chapelle, Sébastien Imperiale.

The objective of this work is to develop a rigorous mathematical and numerical background for the extension and dissemination of elastography imaging modalities, applied to the cardiac setting. The problems treated concern the topics of mathematical modelling, numerical analysis and scientific computing. More precisely, the plan is to define a linearised model for the propagation of elastic waves in the heart, to study approximations of these models and define adapted numerical methods for the discretisation of the resulting partial differential equations.

7.2. Numerical Methods

7.2.1. Effective and energy-preserving time discretization for a general nonlinear poromechanical formulation

Participants: Bruno Burtschell, Dominique Chapelle [correspondant], Philippe Moireau.

In this work, we consider a general nonlinear poromechanical model, formulated based on fundamental thermodynamics principle, suitable for representing the coupling of rapid internal fluid flows with large deformations of the solid, and compatible with a wide class of constitutive behavior. The objective of the present work is to propose for this model a time discretization scheme of the partitioned type, to allow the use of existing time schemes - and possibly separate solvers - for each component of the model, i.e. for the fluid and the solid. To that purpose, we adapt and extend an earlier proposed approach devised for fluid-structure interaction in an Arbitrary Lagrangian-Eulerian framework. We then establish an energy estimate for the resulting time scheme, in a form that is consistent with the underlying energy principle in the poromechanical formulation, up to some numerical dissipation effects and some perturbations that we have carefully identified and assessed. In addition, we provide some numerical illustrations of our numerical strategy with test problems that present typical features of large strains and rapid fluid flows, and also a case of singular transition related to total drainage. An example of challenging application envisioned for this model and associated numerical coupling scheme concerns the perfusion of the heart. This work has resulted in the publication [15].

7.2.2. Delayed feedback control method for calculating space-time periodic solutions of viscoelastic problems

Participants: Ustim Khristenko, Patrick Le Tallec.

We are interested in fast techniques for calculating a periodic solution to viscoelastic evolution problems with a space-time periodic condition. In order to avoid the invertion of very large matrices, such a solution is often computed as an asymptotic limit of the initial value problem with arbitrary initial data. We have developed a control method, accelerating the convergence to the periodic state. The main idea is to modify our problem by introducing a feedback control term, based on a periodicity error.

First, an abstract evolution problem has been studied. From the analytic solution of the modified (controlled) problem, an efficient control has been constructed, optimizing the spectrum of the problem. The proposed control term can be mechanically interpreted, and its efficiency increases with the relaxation time.

In order to confirm numerically the theoretical results, a finite element simulation has been carried out on a full 3D model for a steady rolling of a viscoelastic tyre with periodic sculpture. It has demonstrated that the controlled solution converges indeed faster than the non-controlled one, and that the efficiency of the method increases with the problem's relaxation time, that is when the memory of the underlying problem is large.

7.2.3. Construction and analysis of an adapted spectral finite element method to convective acoustic equations

Participant: Sébastien Imperiale [correspondant].

This work addresses the construction of a non spurious mixed spectral finite element (FE) method to problems in the field of computational aeroacous-tics. Based on a computational scheme for the conservation equations of linear acoustics, the extension towards convected wave propagation is investigated. In aeroacoustic applications, the mean flow effects can have a significant impact on the generated sound field even for smaller Mach numbers. For those convective terms, the initial spectral FE discretization leads to non-physical, spurious solutions. Therefore, a regularization procedure is proposed and qualitatively investigated by means of discrete eigenvalues analysis of the discrete operator in space. A study of convergence and an application of the proposed scheme to simulate the flow induced sound generation in the process of human phonation underlines stability and validity. This work has resulted in the publication [19].

7.2.4. Space/time convergence analysis of a class of conservative schemes for linear wave equations

Participants: Juliette Chabassier [MAGIQUE 3D team], Sébastien Imperiale [correspondant].

This work concerns the space/time convergence analysis of conservative two-steps time discretizations for linear wave equations. Explicit and implicit, second and fourth order schemes are considered, while the space discretization is given and satisfies minimal hypotheses. The convergence analysis is done using energy techniques and holds if the time step is upper-bounded by a quantity depending on space discretization parameters. In addition to showing the convergence for recently introduced fourth order schemes, the novelty of this work consists in the independency of the convergence estimates with respect to the difference between the time step and its greatest admissible value. This work has resulted in the publication [16].

7.3. Inverse Problems

7.3.1. Front observer for data assimilation of electroanatomical mapping data for a numerical atrial model

Participants: Antoine Gérard [Carmen team], Annabelle Collin [Monc team], Jason Bayer [Carmen team], Philippe Moireau [correspondant], Yves Coudière [Carmen team].

The purpose of our work is to personalize an atrial model of the propagation of the action potential, based on electrical catheter data with the help of the data assimilation approach introduced in [Collin & Al, Journal of Computational Physics 2015]. The originality of the approach is to indroduce a Luenberger observer of a surface atrial model of the propagation which can pursue - like in classical Kalman filtering approach - the actual activation front reconstructed from catheter data. Moreover, this approach may account for the breakthrough of new activation fronts at anytime with an additional topological gradient term. In the present work, we adapt this approach to the bilayer surface atrial model of th propagation of action potentials [Labarthe & Al, Europace 2014], and evaluated for the first time on a real patient's dataset. First, the model was geometrically fit to the patient's data. A fiber architecture was added to the geometry. Then an initial electrophysiological state was guessed, and the model was run with the Luenberger filter for some catheter data acquired during a CARTO procedure. All along the simulation, the filter corrects the action potential so as to track CARTO local activation times, while preserving a biophysical behavior. With this technique, we are able to reconstruct smooth activation maps over the whole atrial surfaces. This promising technique may also allow to reconstruct velocity fields and directions, phase map and possibly give information on repolarization. This work results from a collaborative project carried out during a training session at CEMRACS 2016 in Marseille, Luminy. This work has resulted in the publication [28].

7.3.2. Iterative observer-based state and parameter estimation for linear systems

Participant: Atte Aalto [correspondant].

In this work we propose an iterative method for joint state and parameter estimation using measurements on a finite time interval for systems that are backward output stabilizable. Since this time interval is fixed, errors in initial state may have a big impact on the parameter estimate. We propose to use the back and forth nudging (BFN) method for estimating the system's initial state and a Gauss–Newton step between BFN iterations for estimating the system parameters. Taking advantage of results on the optimality of the BFN method, we show that for systems with skew-adjoint generators, the initial state and parameter estimate minimizing an output error cost functional is an attractive fixed point for the proposed method. We treat both linear source estimation and bilinear parameter estimation problems.

7.3.3. Estimation from moments measurements for amyloid depolymerisation

Participants: Aurora Armiento [Mamba team], Marie Doumic [Mamba team], Philippe Moireau [correspondant].

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation, a key feature for amyloid diseases. This study aims at setting this framework problem when the experimental measurements consist in the time-dynamics of a moment of the population (*i.e.* for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering). We propose a general methodology, and we solve the problem theoretically and numerically in the case of a depolymerising system. We then apply our method to experimental data of degrading oligomers, and conclude that smaller aggregates of ovPrP protein should be more stable than larger ones. This has an important biological implication, since it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process. This work has resulted in the publication [14].

7.3.4. Analysis of an observers strategy for initial state reconstruction in unbounded domains

Participants: Antoine Tonnoir [correspondant], Sonia Fliss [Poems team], Sébastien Imperiale, Philippe Moireau.

In this work, we are interested in the problem of recovering a compactly supported initial state of the wave equation in unbounded domain (such as the whole plane, a waveguide...). To this purpose, we assume that the velocity is known in a bounded observation region surrounding the support of the initial state. We consider an iterative algorithm of reconstruction based on back and forth nudging and prove the exponential convergence of this algorithm and its robustness with respect to noisy measures, at the continuous level. We also study the effect of the discretization process on the convergence of the algorithm.

7.4. Experiments and Clinical applications

7.4.1. Characterization of mechanical properties of soft tissues

Participants: Jean-Marc Allain [correspondant], Jean-Sebastien Affagard, Maeva Lopez Poncelas.

Soft tissues - such as skin - have complex mechanical properties: large strains, anisotropy, etc.. Identifying constitutive properties incorporating microstructure effects is very important for applications in medicine (surgery and other therapies) and industry (anti-ageing cosmetics, etc.). This characterization, however, requires complex experiments. We have developed a novel biaxial traction experimental method for mice skin, relying on a sensitivity analysis for determining optimal experimental parameters, including in particular sample size and most informative loading paths. This protocol has already been used on multiple samples, and 3 distinct constitutive laws of increasing complexity have been characterized (Master's internship of Maeva Lopez).

Another originality in our approach is to place our setup under a microscope to monitor the microstructure evolution during the test. These rich measurements allow detailed comparisons of classical models (such as Holzapfel's) with our data.

7.4.2. Non-invasive model-based assessment of passive left-ventricular myocardial stiffness in healthy subjects and in patients with non-ischemic dilated cardiomyopathy

Participant: Radomir Chabiniok [correspondant].

Patient-specific modelling has emerged as a tool for studying heart function, demonstrating the potential to provide non-invasive estimates of tissue passive stiffness. However, reliable use of model-derived stiffness requires sufficient model accuracy and unique estimation of model parameters. In this work we present personalised models of cardiac mechanics, focusing on improving model accuracy, while ensuring unique parametrisation. The influence of principal model uncertainties on accuracy and parameter identifiability was systematically assessed in a group of patients with dilated cardiomyopathy and healthy volunteers. For all cases, we examined three circumferentially symmetric fibre distributions and two epicardial boundary conditions. Our results demonstrated the ability of data-derived boundary conditions to improve model accuracy and highlighted the influence of the assumed fibre distribution on both model fidelity and stiffness estimates. The model personalisation pipeline – based strictly on non-invasive data – produced unique

parameter estimates and satisfactory model errors for all cases, supporting the selected model assumptions. The thorough analysis performed enabled the comparison of passive parameters between volunteers and dilated cardiomyopathy patients, illustrating elevated stiffness in diseased hearts.

7.4.3. Age-related changes in intraventricular kinetic energy: a physiological or pathological adaptation

Participant: Radomir Chabiniok [correspondant].

Aging has important deleterious effects on the cardiovascular system. In this work we sought to compare intraventricular kinetic energy (KE) in healthy subjects of varying ages with subjects with ventricular dysfunction to understand if changes in energetic momentum may predispose individuals to heart failure. Four-dimensional flow MRI was acquired in 35 healthy subjects (age: 1 -67 yr) and 10 patients with left ventricular (LV) dysfunction (age: 28-79 yr). Healthy subjects were divided into age quartiles (1st quartile: 16 yr, 2nd quartile: 17-32 yr, 3rd quartile: 33-48 yr, and 4th quartile: 49 - 64 yr). KE was measured in the LV throughout the cardiac cycle and indexed to ventricular volume. In healthy subjects, two large peaks corresponding to systole and early diastole occurred during the cardiac cycle. A third smaller peak was seen during late diastole in eight adults. Systolic KE (P 0.182) and ejection fraction (P 0.921) were preserved through all age groups. Older adults showed a lower early peak diastolic KE compared with children (P 0.0001) and young adults (P 0.025). Subjects with LV dysfunction had reduced ejection fraction (P 0.001) and compared with older healthy adults exhibited a similar early peak diastolic KE (P 0.142) but with the addition of an elevated KE in diastasis (P 0.029). In healthy individuals, peak diastolic KE progressively decreases with age, whereas systolic peaks remain constant. Peak diastolic KE in the oldest subjects is comparable to those with LV dysfunction. Unique age-related changes in ventricular diastolic energetics might be physiological or herald subclinical pathology. This work has resulted in the publication [24].

7.4.4. Patient-specific computational analysis of ventricular mechanics in pulmonary arterial hypertension

Participant: Martin Genet [correspondant].

Patient-specific biventricular computational models associated with a normal subject and a pulmonary arterial hypertension (PAH) patient were developed to investigate the disease effects on ventricular mechanics. These models were developed using geometry reconstructed from magnetic resonance (MR) images, and constitutive descriptors of passive and active mechanics in cardiac tissues. Model parameter values associated with ventricular mechanical properties and myofiber architecture were obtained by fitting the models with measured pressure–volume loops and circumferential strain calculated from MR images using a hyperelastic warping method. Results show that the peak right ventricle (RV) pressure was substantially higher in the PAH patient (65 mmHg versus 20 mmHg), who also has a significantly reduced ejection fraction (EF) in both ventricles (left ventricle (LV): 39% versus 66% and RV: 18% versus 64%). Peak systolic circumferential strain was comparatively lower in both the left ventricle (LV) and RV free wall (RVFW) of the PAH patient (LV: -6.8% versus -13.2% and RVFW: -2.1% versus -9.4%). Passive stiffness, contractility, and myofiber stress in the PAH patient were all found to be substantially increased in both ventricles, whereas septum wall in the PAH patient possessed a smaller curvature than that in the LV free wall. Simulations using the PAH model revealed an approximately linear relationship between the septum curvature and the transseptal pressure gradient at both early-diastole and end-systole. These findings suggest that PAH can induce LV remodeling, and septum curvature measurements may be useful in quantifying transseptal pressure gradient in PAH patients. This work has resulted in the publication [25].

PARIETAL Project-Team

6. New Results

6.1. Dictionary Learning for Massive Matrix Factorization

Sparse matrix factorization is a popular tool to obtain interpretable data decompositions, which are also effective to perform data completion or denoising. Its applicability to large datasets has been addressed with online and randomized methods, that reduce the complexity in one of the matrix dimension, but not in both of them. In this paper, we tackle very large matrices in both dimensions. We propose a new factoriza-tion method that scales gracefully to terabyte-scale datasets, that could not be processed by previous algorithms in a reasonable amount of time. We demonstrate the efficiency of our approach on massive functional Magnetic Resonance Imaging (fMRI) data, and on matrix completion problems for recommender systems, where we obtain significant speed-ups compared to state-of-the art coordinate descent methods.

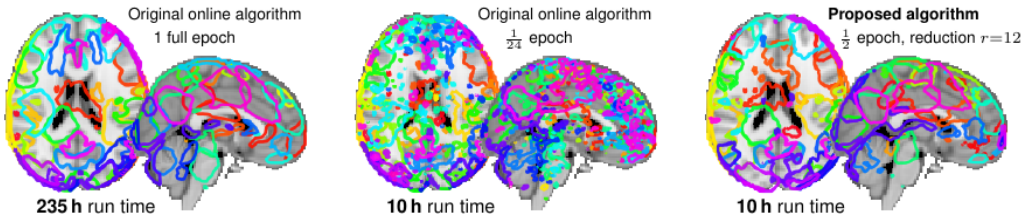


Figure 3. Brain atlases: outlines of each map obtained with dictionary learning. Left: the reference algorithm on the full dataset. Middle: the reference algorithm on a twentieth of the dataset. Right: the proposed algorithm with a similar run time: half the dataset and a compression factor of 9. Compared to a full run of the baseline algorithm, the figure explore two possible strategies to decrease computation time: processing less data (middle), or our approach (right). Our approach achieves a result closer to the gold standard in a given time budget. See [22] for more information.

See Fig. 3 for an illustration and [22] for more information.

6.2. Learning brain regions via large-scale online structured sparse dictionary-learning

We propose a multivariate online dictionary-learning method for obtaining de-compositions of brain images with structured and sparse components (aka atoms). Sparsity is to be understood in the usual sense: the dictionary atoms are constrained to contain mostly zeros. This is imposed via an 1-norm constraint. By "structured", we mean that the atoms are piece-wise smooth and compact, thus making up blobs, as opposed to scattered patterns of activation. We propose to use a Sobolev (Laplacian) penalty to impose this type of structure. Combining the two penalties, we obtain decompositions that properly delineate brain structures from functional images. This non-trivially extends the online dictionary-learning work of Mairal et al. (2010), at the price of only a factor of 2 or 3 on the overall running time. Just like the Mairal et al. (2010) reference method, the online nature of our proposed algorithm allows it to scale to arbitrarily sized datasets. Experiments on brain data show that our proposed method extracts structured and denoised dictionaries that are more intepretable and better capture inter-subject variability in small medium, and large-scale regimes alike, compared to state-of-the-art models.

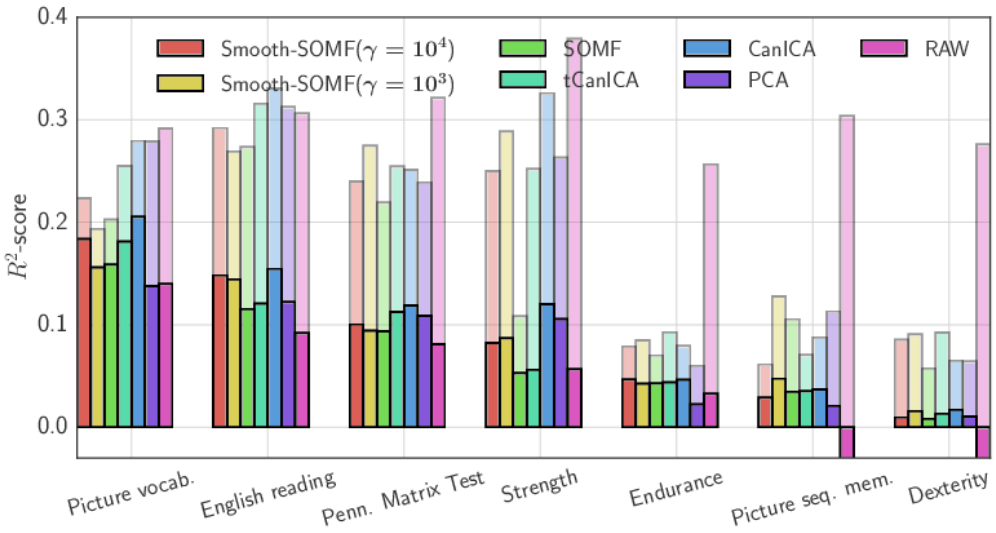


Figure 4. Predicting behavioral variables of the Human Connectome Project dataset using subject-level brain activity maps and various intermediate representations obtained with variants of dictionary learning. Bold bars represent performance on test set while faint bars in the background represent performance on train set. See [19] for more information.

See Fig. 4 for an illustration and [19] for more information.

6.3. Social-sparsity brain decoders: faster spatial sparsity

Spatially-sparse predictors are good models for brain decoding: they give accurate predictions and their weight maps are interpretable as they focus on a small number of regions. However, the state of the art, based on total variation or graph-net, is computationally costly. Here we introduce sparsity in the local neighborhood of each voxel with social-sparsity, a structured shrinkage operator. We find that, on brain imaging classification problems, social-sparsity performs almost as well as total-variation models and better than graph-net, for a fraction of the computational cost. It also very clearly outlines predictive regions. We give details of the model and the algorithm.

See Fig. 5 for an illustration and [32] for more information.

6.4. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example

Resting-state functional Magnetic Resonance Imaging (R-fMRI) holds the promise to reveal functional biomarkers of neuropsychiatric disorders. However, extracting such biomarkers is challenging for complex multi-faceted neuropatholo-gies, such as autism spectrum disorders. Large multi-site datasets increase sample sizes to compensate for this complexity, at the cost of uncontrolled heterogeneity. This heterogeneity raises new challenges, akin to those face in realistic diagnostic applications. Here, we demonstrate the feasibility of inter-site classification of neuropsychiatric status, with an application to the Autism Brain Imaging Data Exchange (ABIDE) database, a large (N=871) multi-site autism dataset. For this purpose, we investigate pipelines that extract the most predictive biomarkers from the data. These R-fMRI pipelines build participant-specific connectomes from functionally-defined brain areas. Connectomes are then compared across participants to

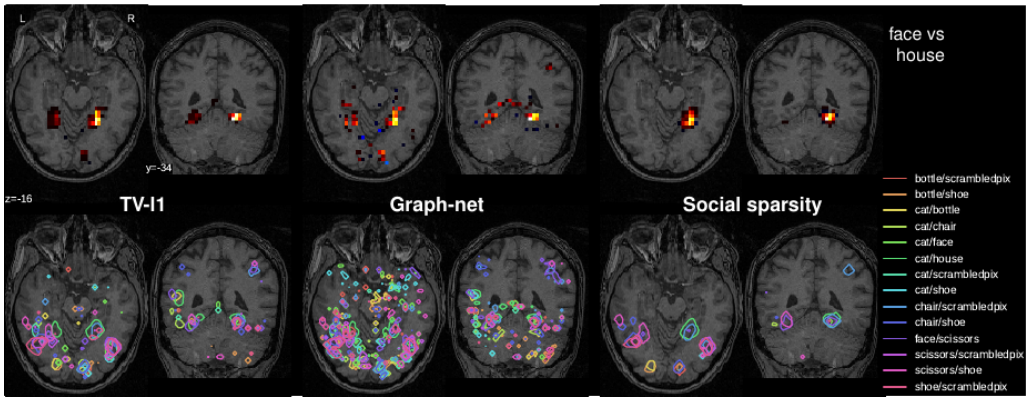


Figure 5. Decoder maps for the object-classification task – Top: weight maps for the face-versus-house task. Overall, the maps segment the right and left parahippocampal place area (PPA), a well-known place-specific regions, although the left PPA is weak in TV-l1, spotty in graph-net, and absent in social sparsity. Bottom: outlines at 0.01 of the other tasks. Beyond the PPA, several known functional regions stand out such as primary or secondary visual areas around the prestriate cortex as well as regions in the lateral occipital cortex, responding to structured objects. Note that the graphnet outlines display scattered small regions even thought the value of the contours is chosen at 0.01, well above numerical noise. See [32] for more information.

learn patterns of connectivity that differentiate typical controls from individuals with autism. We predict this neuropsychiatric status for participants from the same acquisition sites or different, unseen, ones. Good choices of methods for the various steps of the pipeline lead to 67% prediction accuracy on the full ABIDE data, which is significantly better than previously reported results. We perform extensive validation on multiple subsets of the data defined by different inclusion criteria. These enables detailed analysis of the factors contributing to successful connectome-based prediction. First, prediction accuracy improves as we include more subjects, up to the maximum amount of subjects available. Second, the definition of functional brain areas is of paramount importance for biomarker discovery: brain areas extracted from large R-fMRI datasets outperform reference atlases in the classification tasks.

See Fig. 6 for an illustration and [1] for more information.

6.5. Seeing it all: Convolutional network layers map the function of the human visual system

Convolutional networks used for computer vision represent candidate models for the computations performed in mammalian visual systems. We use them as a detailed model of human brain activity during the viewing of natural images by constructing predictive models based on their different layers and BOLD fMRI activations. Analyzing the predictive performance across layers yields characteristic fingerprints for each visual brain region: early visual areas are better described by lower level convolutional net layers and later visual areas by higher level net layers, exhibiting a progression across ventral and dorsal streams. Our predictive model generalizes beyond brain responses to natural images. We illustrate this on two experiments, namely retinotopy and face-place oppositions, by synthesizing brain activity and performing classical brain mapping upon it. The synthesis recovers the activations observed in the corresponding fMRI studies, showing that this deep encoding model captures representations of brain function that are universal across experimental paradigms.

See Fig. 7 for an illustration and [10] for more information.

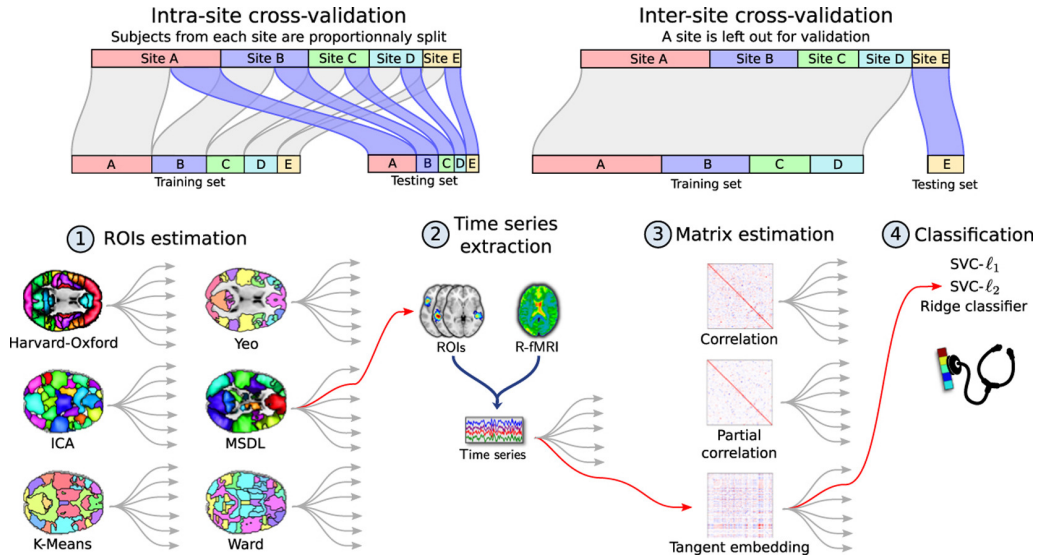


Figure 6. Validation of an fMRI-based pipeline for autism prediction. Several variants are considered for each pipeline step. See [1] for more information.

6.6. Formal Models of the Network Co-occurrence Underlying Mental Operations

Systems neuroscience has identified a set of canonical large-scale networks in humans. These have predominantly been characterized by resting-state analyses of the task-uncon-strained, mind-wandering brain. Their explicit relationship to defined task performance is largely unknown and remains challenging. The present work contributes a multivariate statistical learning approach that can extract the major brain networks and quantify their configuration during various psychological tasks. The method is validated in two extensive datasets (n = 500 and n = 81) by model-based generation of synthetic activity maps from recombination of shared network topographies. To study a use case, we formally revisited the poorly understood difference between neural activity underlying idling versus goal-directed behavior. We demonstrate that task-specific neural activity patterns can be explained by plausible combinations of resting-state networks. The possibility of decomposing a mental task into the relative contributions of major brain networks, the "network co-occurrence architecture" of a given task, opens an alternative access to the neural substrates of human cognition.

See Fig. 8 for an illustration and [6] for more information.

6.7. Transmodal Learning of Functional Networks for Alzheimer's Disease Prediction

Functional connectivity describes neural activity from resting-state functional magnetic resonance imaging (rs-fMRI). This noninvasive modality is a promising imaging biomarker of neurodegenerative diseases, such as Alzheimer's disease (AD), where the connectome can be an indicator to assess and to understand the pathology. However, it only provides noisy measurements of brain activity. As a consequence, it has shown fairly limited discrimination power on clinical groups. So far, the reference functional marker of AD is the fluorodeoxyglucose positron emission tomography (FDG-PET). It gives a reliable quantification of metabolic activity, but it is costly and invasive. Here, our goal is to analyze AD populations solely based on rs-fMRI, as functional connectivity is correlated to metabolism. We introduce transmodal learning: leveraging a prior

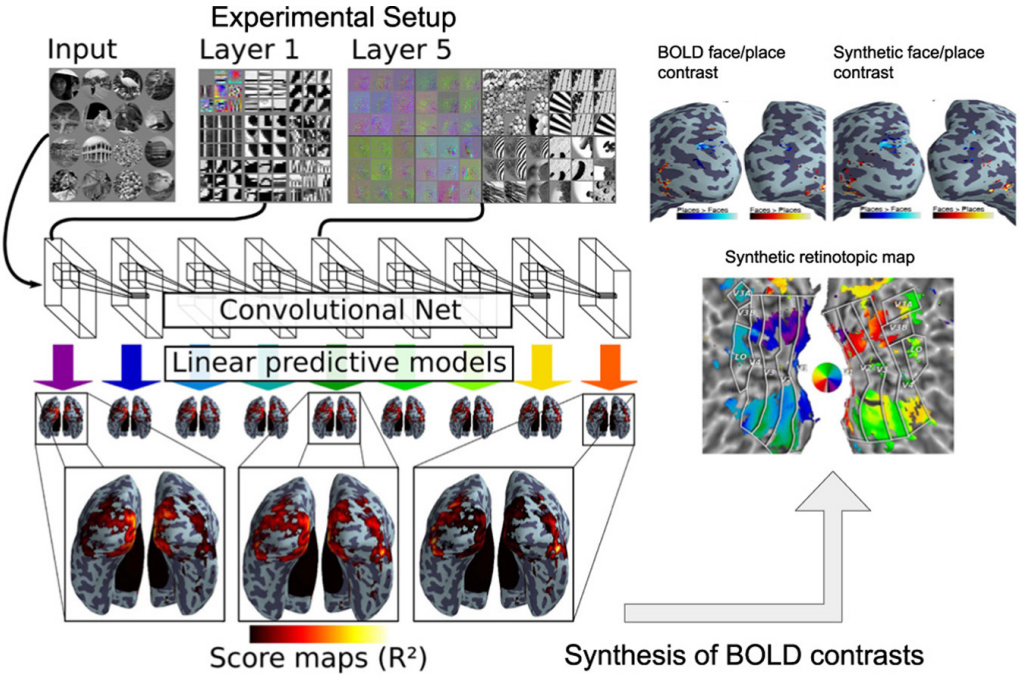


Figure 7. Overview of the vision mapping experiment: Convolutional network image representations of different layer depth explain brain activity throughout the full ventral visual stream. This mapping follows the known hierarchical organisation. Results from both static images and video stimuli. A model of brain activity for the full brain, based on the convolutional network, can synthesize brain maps for other visual experiments. Only deep models can reproduce observed BOLD activity. See [10] for more information.

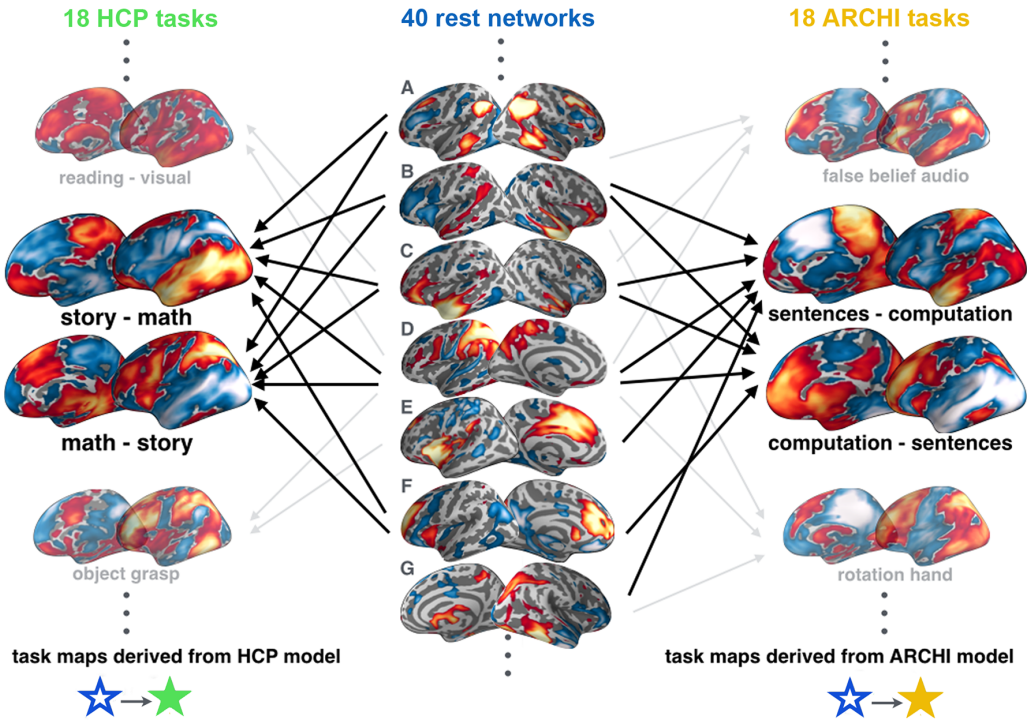


Figure 8. Task-rest correspondence: Reconstructing two similar tasks from two different datasets based on the same resting networks. 40 sparse PCA networks were discovered from the same rest data and used for feature engineering as a basis for classificationof 18 psychological tasks from HCP (left) and from ARCHI (right). Middle column: Examples of resting-state networks derived from decomposing rest data using sparse PCA. Networks B and C might be related to semantics processing in the anterior temporal lobe, network D covers extended parts of the parietal cortex, while networks E and F appear to be variants of the so-called "salience" network. Left/Right column: Examples of task-specific neural activity generated from network co-occurrence models of the HCP/ARCHI task batteries. Arrows: A diagnostic subanalysis indicated what rest networks were automatically ranked top-five in distinguishing a given task from the respective 17 other tasks. Although the experimental tasks in the HCP and ARCHI repositories, "story versus math" and "sentences versus computation" were the most similar cognitive contrasts in both datasets. For these four experimental conditions the model-derived task maps are highly similar. Consequently, two independent classification problems in two independent datasets with a six-fold difference in sample size resulted in two independent explicit models that, nevertheless, generated comparable task-specific maps. This indicated that network co-occurrence modeling indeed captures genuine aspects of neurobiology rather than arbitrary discriminatory aspects of the data. See [6] for more information.

from one modality to improve results of another modality on different subjects. A metabolic prior is learned from an independent FDG-PET dataset to improve functional connectivity-based prediction of AD. The prior acts as a regularization of connectivity learning and improves the estimation of discriminative patterns from distinct rs-fMRI datasets. Our approach is a two-stage classification strategy that combines several seed-based connectivity maps to cover a large number of functional networks that identify AD physiopathology. Experimental results show that our transmodal approach increases classification accuracy compared to pure rs-fMRI approaches, without resorting to additional invasive acquisitions. The method successfully recovers brain regions known to be impacted by the disease.

6.8. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines

Decoding, ie prediction from brain images or signals, calls for empirical evaluation of its predictive power. Such evaluation is achieved via cross-validation, a method also used to tune decoders' hyper-parameters. This paper is a review on cross-validation procedures for decoding in neuroimaging. It includes a didactic overview of the relevant theoretical considerations. Practical aspects are highlighted with an extensive empirical study of the common decoders in within-and across-subject predictions, on multiple datasets –anatomical and functional MRI and MEG– and simulations. Theory and experiments outline that the popular " leave-one-out " strategy leads to unstable and biased estimates, and a repeated random splits method should be preferred. Experiments outline the large error bars of cross-validation in neuroimaging settings: typical confidence intervals of 10%. Nested cross-validation can tune decoders' parameters while avoiding circularity bias. However we find that it can be more favorable to use sane defaults, in particular for non-sparse decoders.

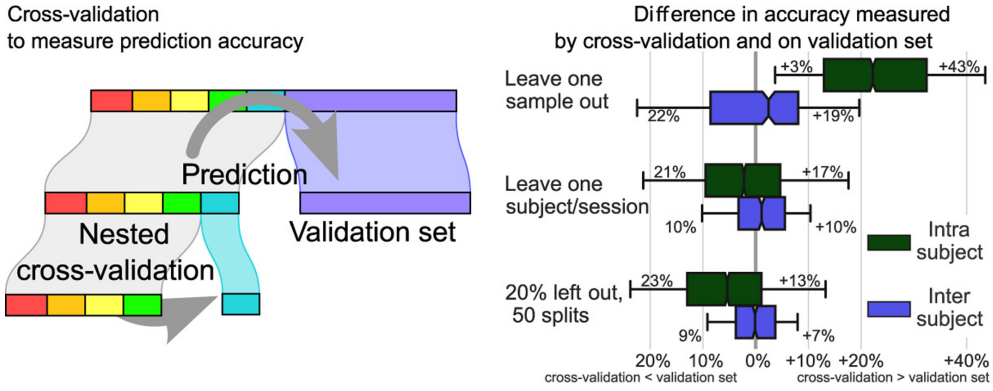


Figure 9. (Left) Illustration of the nested cross-validation principle. (Right) Typical cross-validated accuracy result: leave-one-out cross validation, when applied to imaging data, yields to optimistic bias (top) when used on dependent data, and in other cases leads to estimated with inflated variance. See [16] for more information.

See Fig. 9 for an illustration and [16] for more information.

6.9. A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging

Collecting the maximal amount of information in a given scanning time is a major concern in Magnetic Resonance Imaging (MRI) to speed up image acquisition. The hardware constraints (gradient magnitude, slew rate, ...), physical distortions (e.g., off-resonance effects) and sampling theorems (Shannon, compressed sensing) must be taken into account simultaneously, which makes this problem extremely challenging. To date,

the main approach to design gradient waveform has consisted of selecting an initial shape (e.g. spiral, radial lines, ...) and then traversing it as fast as possible using optimal control. In this paper, we propose an alternative solution which first consists of defining a desired parameterization of the trajectory and then of optimizing for minimal deviation of the sampling points within gradient constraints. This method has various advantages. First, it better preserves the density of the input curve which is critical in sampling theory. Second, it allows to smooth high curvature areas making the acquisition time shorter in some cases. Third, it can be used both in the Shannon and CS sampling theories. Last, the optimized trajectory is computed as the solution of an efficient iterative algorithm based on convex programming. For piecewise linear trajectories, as compared to optimal control reparameterization, our approach generates a gain in scanning time of 10% in echo planar imaging while improving image quality in terms of signal-to-noise ratio (SNR) by more than 6 dB. We also investigate original trajectories relying on traveling salesman problem solutions. In this context, the sampling patterns obtained using the proposed projection algorithm are shown to provide significantly better reconstructions (more than 6 dB) while lasting the same scanning time.

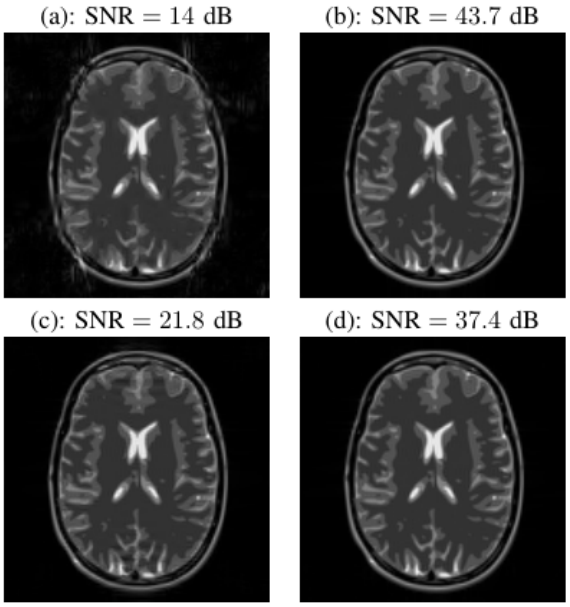


Figure 10. Reconstructed images from data collected along EPI-like trajectories. (a)-(b): Reconstruction results from the optimally reparameterized EPI readout. (c)-(d): Reconstructed results from data collected using the projected EPI trajectories. See [9] for more information.

See Fig. 10 for an illustration and [9] for more information.

6.10. Impact of perceptual learning on resting-state fMRI connectivity: A supervised classification study

Perceptual learning sculpts ongoing brain activity. This finding has been observed by statistically comparing the functional connectivity (FC) patterns computed from resting-state functional MRI (rs-fMRI) data recorded before and after intensive training to a visual attention task. Hence, functional connectivity serves a dynamic role in brain function, supporting the consolidation of previous experience. Following this line of research, we trained three groups of individuals to a visual discrimination task during a magneto-encephalography (MEG) experiment. The same individuals were then scanned in rs-fMRI. Here, in a supervised classification

framework, we demonstrate that FC metrics computed on rs-fMRI data are able to predict the type of training the participants received. On top of that, we show that the prediction accuracies based on tangent embedding FC measure outperform those based on our recently developed multivariate wavelet-based Hurst exponent estimator, which captures low frequency fluctuations in ongoing brain activity too.

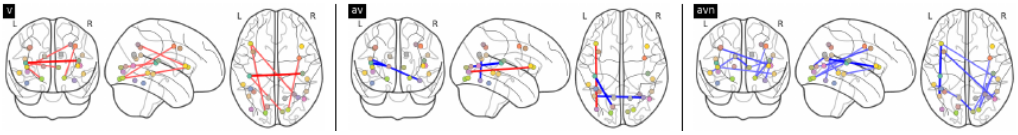


Figure 11. Statistical significant functional interactions (positive and negative values are color coded in red and blue, respectively) within each group of individuals (V: purely visual traing, AV: audio-visual training and AVn: unmatched audio-visual), Bonferroni-corrected for multiple comparisons at $\alpha = 0.05$. *See [24] for more information.*

See Fig. 11 for an illustration and [24] for more information.

XPOP Team

7. New Results

7.1. Identifiability in mixed effects models

We considered the question of model identifiability within the context of nonlinear mixed effects models. Although there has been extensive research in the area of fixed effects models, much less attention has been paid to random effects models.

In this context we distinguish between theoretical identifiability, in which different parameter values lead to non-identical probability distributions, structural identifiability which concerns the algebraic properties of the structural model, and practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise.

We have explored a number of pharmacokinetic models which are known to be non-identifiable at an individual level but can become identifiable at the population level if a number of specific assumptions on the probabilistic model hold. Essentially if the probabilistic models are different, even though the structural models are non-identifiable, then they will lead to different likelihoods. The findings are supported through simulations.

7.2. Enhanced Method for Diagnosing Pharmacometric Models

For nonlinear mixed-effects pharmacometric models, diagnostic approaches often rely on individual parameters, also called empirical Bayes estimates (EBEs), estimated through maximizing conditional distributions. When individual data are sparse, the distribution of EBEs can "shrink" towards the same population value, and as a direct consequence, resulting diagnostics can be misleading.

Instead of maximizing each individual conditional distribution of individual parameters, we propose to randomly sample them in order to obtain values better spread out over the marginal distribution of individual parameters.

We have evaluated, through diagnostic plots and statistical tests, hypothesis related to the distribution of the individual parameters and shown that the proposed method leads to more reliable results than using the EBEs. In particular, diagnostic plots are more meaningful, the rate of type I error is correctly controlled and its power increases when the degree of misspecification increases. An application to the warfarin pharmacokinetic data confirms the interest of the approach for practical applications.

7.3. A Shrinkage-Thresholding Metropolis Adjusted Langevin Algorithm for Bayesian Variable Selection

We have introduced a new Markov Chain Monte Carlo method for Bayesian variable selection in high dimensional settings. The algorithm is a Hastings-Metropolis sampler with a proposal mechanism which combines a Metropolis Adjusted Langevin (MALA) step to propose local moves associated with a shrinkage-thresholding step allowing to propose new models.

The geometric ergodicity of this new trans-dimensional Markov Chain Monte Carlo sampler was established. An extensive numerical experiment, on simulated and real data, illustrates the performance of the proposed algorithm in comparison with some more classical trans-dimensional algorithms

7.4. Maximum likelihood estimation of a low-order building model

Our objective was to investigate the accuracy of the estimates learned with an open loop model of a building whereas the data is actually collected in closed loop, which corresponds to the true exploitation of buildings. We have proposed a simple model based on an equivalent RC network whose parameters are physically interpretable. We also described the maximum likelihood estimation of these parameters by the EM algorithm, and derived their statistical properties.

The numerical experiments clearly show the potential of the method, in terms of accuracy and robustness. We have emphasized the fact that the estimations are linked to the generating process for the observations, which includes the command system. For instance, the features of the building are correctly estimated if there is a significant gap between the heating and cooling setpoint.

7.5. LP-convergence of a Girsanov theorem based particle filter

We have analyzed the LP-convergence of a previously proposed Girsanov theorem based particle filter for discretely observed stochastic differential equation (SDE) models. We proved the convergence of the algorithm with the number of particles tending to infinity by requiring a moment condition and a step-wise initial condition boundedness for the stochastic exponential process giving the likelihood ratio of the SDEs. The practical implications of the condition are illustrated with an Ornstein–Uhlenbeck model and with a non-linear Bene's model.

7.6. Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding

In the random coefficients binary choice model, a binary variable equals 1 iff an index $X^T\beta$ is positive. The vectors X and β are independent and belong to the sphere S^{d-1} in R^d. We have proven lower bounds on the minimax risk for estimation of the density f_β over Besov bodies where the loss is a power of the $L^p(S^{d-1})$ norm for $1 \leq p \leq \infty$. We have shown that a hard thresholding estimator based on a needlet expansion with data-driven thresholds achieves these lower bounds up to logarithmic factors.

INFINE Project-Team

6. New Results

6.1. Online Social Networks (OSN)

Community detection; bandit algorithms; privacy preservation; reward mechanisms

6.1.1. Capacity of Information Processing Systems

Participants: Laurent Massoulié, Kuang Xu.

We propose and analyze a family of information processing systems, where a finite set of experts or servers are employed to extract information about a stream of incoming jobs. Each job is associated with a hidden label drawn from some prior distribution. An inspection by an expert produces a noisy outcome that depends both on the job's hidden label and the type of the expert, and occupies the expert for a finite time duration. A decision maker's task is to dynamically assign inspections so that the resulting outcomes can be used to accurately recover the labels of all jobs, while keeping the system stable. Among our chief motivations are applications in crowd-sourcing, diagnostics, and experiment designs, where one wishes to efficiently learn the nature of a large number of items, using a finite pool of computational resources or human agents. We focus on the capacity of such an information processing system. Given a level of accuracy guarantee, we ask how many experts are needed in order to stabilize the system, and through what inspection architecture. Our main result provides an adaptive inspection policy that is asymptotically optimal in the following sense: the ratio between the required number of experts under our policy and the theoretical optimal converges to one, as the probability of error in label recovery tends to zero.

This work was accepted and presented under the title "On the capacity of information processing systems" at the COLT 2016 conference.

6.2. Spontaneous Wireless Networks and Internet of Things

internet of things; wireless sensor networks; dissemination; resource management

6.2.1. Platform Design for the Internet of Things

Participants: Emmanuel Baccelli, Cedric Adjih, Oliver Hahm, Francisco Acosta, Hauke Petersen.

Within this activity, we have further developed the platforms we champion for the Internet of Things: the open source operating system RIOT on one hand, and open-access IoT-lab testbeds on the other hand. RIOT now aggregates open source contributions from 130+ people (and counting) from all over the world, coming both from academia and from industry, and received financial backing from top companies including Cisco and Google. We further developed RIOT for low-cost mobile robots and received the Best Demo Award at the ACM EWSN'16 conference for our work on this topic. As steering RIOT community members, we also participated in the prestigious Internet Architecture Board (IAB) workshop on IoT Software Updates, a hot and essential topic for the future of Internet of Things. The year culminated in this domain with the successful organization of the first RIOT Summit in Berlin, where 100+ participants from all over the world, from industry, academia as well as hackers/makers involved in RIOT gathered to discuss various aspects of the future of RIOT and open source IoT software. In addition, 2016, at the site of Saclay, one of the testbeds from FIT IoT-LAB was opened: the platform of Saclay includes more than 300 IoT nodes (175 A8-M3, 12 M3, 120 WSN430, some Arduinos and some SAMR21-xpro). In parallel, the platform from Freie Universitat Berlin also joined the OneLab/FIT IoT-LAB testbed federation.

6.2.2. Energy-Efficient Communication Protocols for the Internet of Things

Participants: Oliver Hahm, Emmanuel Baccelli, Cedric Adjih, Matthias Waehlisch, Thomas Schmidt.

Within this activity, we have designed distributed algorithms providing improved trade-off between content availability and energy efficiency (which plays a crucial role). The approach we developed leverages distributed caching for IoT content, based on an information-centric networking paradigm. We extended the NDN protocol with a variety of caching and replacement strategies, and we analyzed alternative approaches for extending NDN to accommodate such IoT use cases. Based on extensive experiments on real IoT hardware in a network gathering hundreds of nodes, we demonstrate these caching strategies can bring 90% reduction in energy consumption while maintaining IoT content availability above 90%. This work was published in IEEE Globecom'16 workshop on Named Data Networks for Challenged Communication Environments.

We also have designed new mechanisms to jointly exploit ICN communication patterns and dynamically optimize the use of TSCH (Time Slotted Channel Hopping), a wireless link layer technology increasingly popular in the IoT. Through a series of experiments on FIT IoT-LAB interconnecting typical IoT hardware, we find that our proposal is fully robust against wireless interference, and almost halves the energy consumed for transmission when compared to CSMA. Most importantly, our adaptive scheduling prevents the time-slotted MAC layer from sacrificing throughput and delay. Our work on ICN and on TSCH was published at NTMS'16, at ACM ICN'16, and in Proceedings of the IEEE.

6.2.3. Standards for Spontaneous Wireless Networks

Participant: Emmanuel Baccelli.

Within this activity, we have contributed to new network protocol standards for spontaneous wireless networking, applied to ad hoc networks and the Internet of Things. In particular, collaborating with Fraunhofer, we have published RFC 7779, standardizing Directional Airtime Metric (DAT), a new wireless metric standard targeting wireless mesh networks. Furthermore, collaborating with ARM and Sigma Designs, we published RFC 7733, which provides guidance in the configuration and use of protocols from the RPL protocol suite to implement the features required for control in building and home environments. In collaboration with various industrial partners, with have also published a number of other Internet drafts, including an analysis of the characteristics of multi-hop ad hoc wireless communication between interfaces in the context of IP networks, and an analysis of the challenges of information-centric networking in the Internet of Things.

6.2.4. Spatio-Temporal Predictability of Cellular Data Traffic

Participants: Guangshuo Chen, Aline Carneiro Viana, Marco Fiore, Sahar Hoteit, Carlos Sarraute.

The ability to foresee the data traffic activity of subscribers opens new opportunities to reshape mobile network management and services. In this work, we leverage two large-scale real-world datasets collected by a major mobile carrier in Mexico to study how predictable are the cellular data traffic demands generated by individual users. We focus on the predictability of mobile traffic consumption patterns in isolation. Our results show that it is possible to anticipate the individual demand with a typical accuracy of 85%, and reveal that this percentage is consistent across all user types. Despite the heterogeneity in usage patterns of users, we also find a lack of significant variability in predictability when considering demographic factors or different mobility or mobile service usage. We also analyze the joint predictability of the traffic demands and mobility patterns. We find that the two dimensions are correlated, which improves the predictability upper bound to 90%on average. This first work is in submission in an international conference.

6.2.5. Completion of Sparse Call Detail Records for Mobility Analysis

Participants: Guangshuo Chen, Aline Carneiro Viana, Marco Fiore, Sahar Hoteit.

Call Detail Records (CDRs) have been widely used in the last decades for studying different aspects of human mobility. The accuracy of CDRs strongly depends on the user-network interaction frequency: hence, the temporal and spatial sparsity that typically characterize CDR can introduce a bias in the mobility analysis. In this work, we evaluate the bias induced by the use of CDRs for inferring important locations of mobile subscribers, as well as their complete trajectories. Besides, we propose a novel technique for estimating real human trajectories from sparse CDRs. Compared to previous solutions in the literature, our proposed technique reduces the error between real and estimated human trajectories and at the same time shortens the temporal

period where users' locations remain undefined. This work has been published as an invited paper at the ACM CHANTS 2016 workshop in conjunction with ACM MobiCom 2016. Related to CDRs, we have also investigated whether the information of user's instantaneous whereabouts provided by CDRs enables us to estimate positions over longer time spans. Our results confirm that CDRs ensure a good estimation of radii of gyration and important locations, yet they lose some location information. Most importantly, we show that temporal completion of CDRs is straightforward and efficient: thanks to the fact that they remain fairly static before and after mobile communication activities, the majority of users' locations over time can be accurately inferred from CDRs. Finally, we observe the importance of user's context, i.e.,of the size of the current network cell, on the quality of the CDR temporal completion. This work is in submission in an international conference. Finally, driven by real-world data across a large population, we propose two approaches as the refinement of the legacy solution, which complete CDR data adaptively according to the information of users and activities. Our proposed methods outperform the legacy solution in terms of the combination of accuracy and temporal coverage. Besides, our work reveals the important factors to the data completion. This paper has been accepted for publication at the IEEE DAWM workshop in conjunction with IEEE Percom 2017.

6.2.6. Completion of Sparse Call Detail Records for Mobility Analysis

Participants: Panagiota Katsikouli, Aline Carneiro Viana, Marco Fiore, Alessandro Nordio, Alberto Tarable.

The increasing usage of smart devices and location-tracking systems has made it possible to study and understand the behaviour of users as well as human mobility at an unprecedented scale. The insights of such studies can help improve many aspects of our everyday lives, from road network infrastructure to mobile network quality of service. Human mobility is repetitive and regular.In addition to our tendency to revisit the same locations, those visits happen with relevant temporal regularity, where each visited location has been assigned with an ID. The daily interaction with our smart devices, such as smartphones, results in collecting fine grained information on our activities and whereabouts. This information can be used to detect and analyze the routinary behaviour of humans but also to discover interests, preferences and hidden patterns of mobility. However, frequent recording of data tends to quickly drain the battery of the smartphone. A natural alternative is to sample the collected data. Maintaining a summary or sample as close to the original collected data as possible is the key challenge. Deciding what constitutes a representative sample depends on the type of information we wish to maintain from the data collected. In this work, we wish to sparsely sample mobility traces of GPS data with the goal to reconstruct the movement of the users both in space and time at the desired granularity. An ideal sample would allow us to reconstruct the traces in such a way that we preserve the frequency of visits and the time spent to the various locations. Therefore, the problem we tackle here is to *sparsely sample the mobility trace of a user with the goal to reconstruct her complete trace in space and time*. This is an on-going work and will be submitted to an international conference in the next months.

6.3. Resource and Traffic Management

Traffic offloading; infrastructure deployment; opportunistic routing; traffic modeling; intermittently connected networks.

6.3.1. Utility Optimization Approach to Network Cache Design

Participants: Mostafa Dehghan, Laurent Massoulié, Don Towsley, Daniel Menasche, Y.c. Tay.

In any caching system, the admission and eviction policies determine which contents are added and removed from a cache when a miss occurs. Usually, these policies are devised so as to mitigate staleness and increase the hit probability. Nonetheless, the utility of having a high hit probability can vary across contents. This occurs, for instance, when service level agreements must be met, or if certain contents are more difficult to obtain than others. In this paper, we propose utility-driven caching, where we associate with each content a utility, which is a function of the corresponding content hit probability. We formulate optimization problems where the objectives are to maximize the sum of utilities over all contents. These problems differ according to the stringency of the cache capacity constraint. Our framework enables us to reverse engineer classical replacement policies such as LRU and FIFO, by computing the utility functions that they maximize. We also develop online algorithms that can be used by service providers to implement various caching policies based on arbitrary utility functions.

This work was published and presented at the IEEE Infocom 2016 conference as ”A Utility Optimization Approach to Network Cache Design”.

AVIZ Project-Team

7. New Results

7.1. Swarm User Interfaces

Participants: Mathieu Le Goc [correspondant], Lawrence Kim, Ali Parsaei, Jean-Daniel Fekete, Pierre Dragicevic, Sean Follmer.

We introduce swarm user interfaces (Fig 3), a new class of human-computer interfaces comprised of many autonomous robots that handle both display and interaction. We describe the design of Zooids, a hardware and software system: a small wheel-propelled robot with position and touch sensing capabilities that can be freely arranged and repositioned on any horizontal surface, both through user manipulation and computer control. Zooids is an open-source open-hardware platform for developing tabletop swarm interfaces. We illustrate the potential of tabletop swarm user interfaces through a set of application scenarios developed with Zooids, and discuss general design considerations unique to swarm user interfaces.

More on the project Web page: http://www.aviz.fr/swarmui.

7.2. A Systematic Review of Experimental Studies on Data Glyphs

Participants: Johannes Fuchs, Petra Isenberg [correspondant], Anastasia Bezerianos, Daniel Keim.

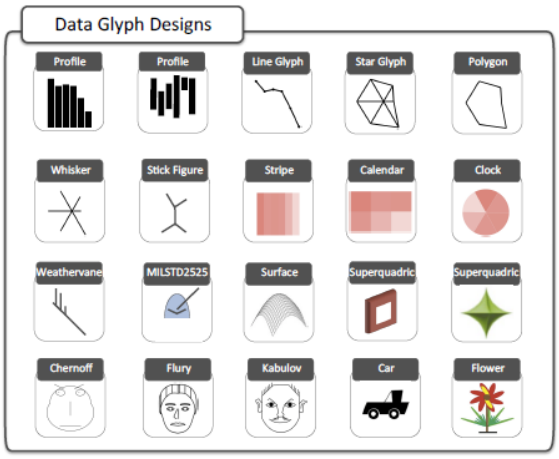


Figure 9. Overview of the glyphs reviewed in the study.

We systematically reviewed 64 user-study papers on data glyphs to help researchers and practitioners gain an informed understanding of tradeoffs in the glyph design space. The glyphs we considered were individual representations of multi-dimensional data points, often meant to be shown in small-multiple settings. Over the past 60 years many different glyph designs were proposed and many of these designs have been subjected to perceptual or comparative evaluations. Yet, a systematic overview of the types of glyphs and design variations tested, the tasks under which they were analyzed, or even the study goals and results did not yet exist. We provide such an overview by systematically sampling and tabulating the literature on data glyph studies, listing their designs, questions, data, and tasks. In addition we present a concise overview of the types of glyphs and their design characteristics analyzed by researchers in the past, and a synthesis of the study results. Based on our meta analysis of all results we further contribute a set of design implications and a discussion on open research directions.

7.3. Towards an Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration

Participants: David López, Lora Oehlberg, Candemir Doger, Tobias Isenberg [correspondant].

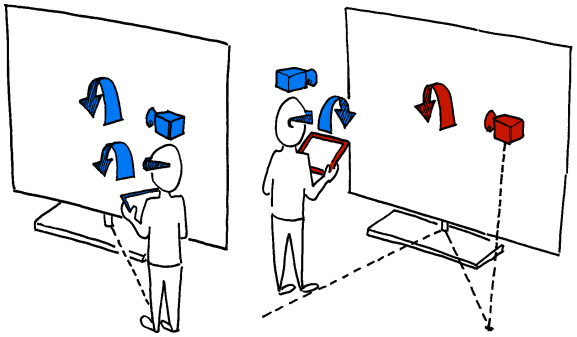


Figure 10. Illustration of the problem of mobility within a virtual environment, while interacting with a view on a tablet.

We discuss touch-based navigation of 3D visualizations in a combined monoscopic and stereoscopic viewing environment. We identify a set of interaction modes, and a workflow that helps users transition between these modes to improve their interaction experience. In our discussion we analyze, in particular, the control-display space mapping between the different reference frames of the stereoscopic and monoscopic displays. We show how this mapping supports interactive data exploration, but may also lead to conflicts between the stereoscopic and monoscopic views due to users' movement in space; we resolve these problems through synchronization. To support our discussion, we present results from an exploratory observational evaluation with domain experts in fluid mechanics and structural biology. These experts explored domain-specific datasets using variations of a system that embodies the interaction modes and workflows; we report on their interactions and qualitative feedback on the system and its workflow.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Lopez2016TUM.

7.4. CAST: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds

Participants: Lingyun Yu, Konstantinos Efstathiou, Petra Isenberg, Tobias Isenberg [correspondant].

We present a family of three interactive Context-Aware Selection Techniques (CAST) for the analysis of large 3D particle datasets. For these datasets, spatial selection is an essential prerequisite to many other analysis tasks. Traditionally, such interactive target selection has been particularly challenging when the data subsets of interest were implicitly defined in the form of complicated structures of thousands of particles. Our new techniques SpaceCast, TraceCast, and PointCast improve usability and speed of spatial selection in point clouds through novel context-aware algorithms. They are able to infer a user's subtle selection intention from gestural input, can deal with complex situations such as partially occluded point clusters or multiple cluster layers, and can all be fine-tuned after the selection interaction has been completed. Together, they provide an effective and efficient tool set for the fast exploratory analysis of large datasets. In addition to presenting Cast, we report on a formal user study that compares our new techniques not only to each other but also to existing state-of-the-art selection methods. Our results show that Cast family members are virtually always faster than existing methods without tradeoffs in accuracy. In addition, qualitative feedback shows that PointCast and TraceCast were strongly favored by our participants for intuitiveness and efficiency.

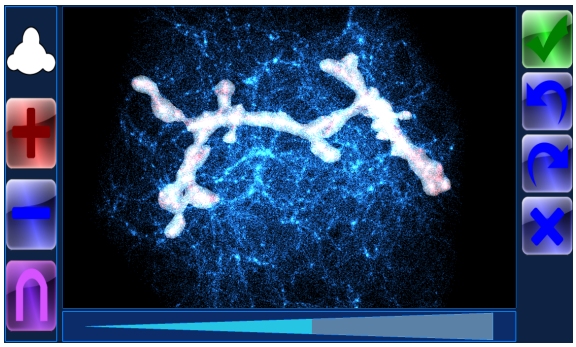


Figure 11. Illustration of the complexity metrics used in the study.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Yu2016CEE.

7.5. Hybrid Tactile/Tangible Interaction for 3D Data Exploration

Participants: Lonni Besançon [correspondant], Paul Issartel, Mehdi Ammi, Tobias Isenberg.

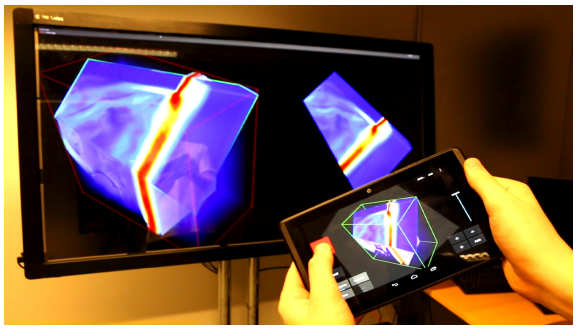


Figure 12. Picture of the hybrid interaction system.

We present the design and evaluation of an interface that combines tactile and tangible paradigms for 3D visualization. While studies have demonstrated that both tactile and tangible input can be efficient for a subset of 3D manipulation tasks, we reflect here on the possibility to combine the two complementary input types. Based on a field study and follow-up interviews, we present a conceptual framework of the use of these different interaction modalities for visualization both separately and combined—focusing on free exploration as well as precise control. We present a prototypical application of a subset of these combined mappings for fluid dynamics data visualization using a portable, position-aware device which offers both tactile input and tangible sensing. We evaluate our approach with domain experts and report on their qualitative feedback.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/HybridInteraction/HybridInteraction.html.

7.6. A Tangible Volume for Portable 3D Interaction

Participants: Paul Issartel, Lonni Besançon [correspondant], Tobias Isenberg, Mehdi Ammi.

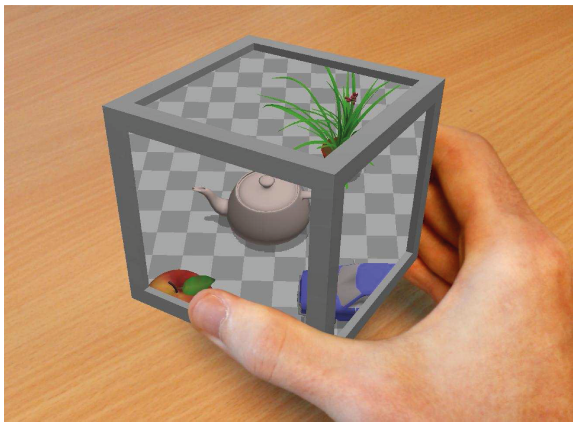


Figure 13. Image of the Cube.

We present a new approach to achieve tangible object manipulation with a single, fully portable and self-contained device. Our solution is based on the concept of a "tangible volume". We turn a tangible object into a handheld fish-tank display. The tangible volume represents a volume of space that can be freely manipulated within a virtual scene. This volume can be positioned onto virtual objects to directly grasp them, and to manipulate them in 3D space. We investigate this concept through two user studies. The first study evaluates the intuitiveness of using a tangible volume for grasping and manipulating virtual objects. The second study evaluates the effects of the limited field of view on spatial awareness. Finally, we present a generalization of this concept to other forms of interaction through the surface of the volume.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/TangibleCube/TangibleCube.html.

7.7. Preference Between Allocentric and Egocentric 3D Manipulation in a Locally Coupled Configuration

Participants: Paul Issartel, Lonni Besançon [correspondant], Steven Franconeri.

We study user preference between allocentric and egocentric 3D manipulation on mobile devices, in a configuration where the motion of the device is applied to an object displayed on the device itself. We first evaluate this preference for translations and for rotations alone, then for full 6-DOF manipulation. We also investigate the role of contextual cues by performing this experiment in different 3D scenes. Finally, we look at the specific influence of each manipulation axis. Our results provide guidelines to help interface designers select an appropriate default mapping in this locally coupled configuration.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/Mappings/Mappings.html.

7.8. Embedded Data Representations

Participants: Wesley Willett, Yvonne Jansen, Pierre Dragicevic [correspondant].

We introduces *embedded data representations* as the use of visual and physical representations of data that are deeply integrated with the physical spaces, objects, and entities to which the data refers [16]. Technologies like lightweight wireless displays, mixed reality hardware, and autonomous vehicles are making it increasingly easier to display data in-context. While researchers and artists have already begun to create embedded data representations, the benefits, trade-offs, and even the language necessary to describe and compare these approaches remain unexplored.

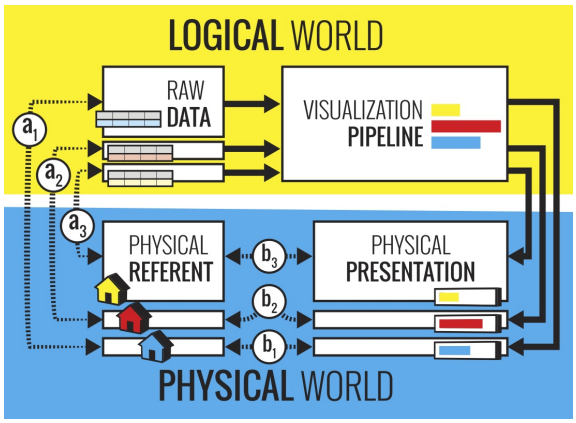


Figure 14. Conceptual model for situated and embedded data representations.

In this paper, we formalize the notion of physical data referents – the real-world entities and spaces to which data corresponds – and examine the relationship between referents and the visual and physical representations of their data. We differentiate situated representations, which display data in proximity to data referents, and embedded representations, which display data so that it spatially coincides with data referents. Drawing on examples from visualization, ubiquitous computing, and art, we explore the role of spatial indirection, scale, and interaction for embedded representations. We also examine the tradeoffs between non-situated, situated, and embedded data displays, including both visualizations and physicalizations. Based on our observations, we identify a variety of design challenges for embedded data representation, and suggest opportunities for future research and applications

7.9. Space-Time Cube Framework

Participants: Benjamin Bach, Pierre Dragicevic [correspondant], Dominique Archambault, Christophe Hurter, Sheelagh Carpendale.

We presented a descriptive model for visualizations of temporal data based on a generalized space-time cube framework [1]. Visualizations are described as operations on a conceptual space-time cube, which transform the cube's 3D shape into readable 2D visualizations. Operations include: extracting subparts of the cube, flattening it across space or time, or transforming the cube's geometry and content. We introduced a taxonomy of elementary space-time cube operations and explained how these operations can be combined and parameterized.

The generalized space-time cube has two properties: a) it is purely conceptual without the need to be implemented, and b) it applies to all datasets that can be represented in two dimensions plus time (e.g., geospatial, videos, networks, multivariate data). The proper choice of space-time cube operations depends on many factors, e.g., density or sparsity of a cube, hence we proposed a characterization of structures within space-time cubes, which allowed us to discuss strengths and limitations of operations. We also reviewed interactive systems that support multiple operations, allowing a user to customize his view on the data. With this framework, we hope to facilitate the description, criticism and comparison of temporal data visualizations, as well as encourage the exploration of new techniques and systems.

More on the project Web page: spacetimecubevis.com.

7.10. The Attraction Effect in Information Visualization

Participants: Evanthia Dimara [correspondant], Anastasia Bezerianos, Pierre Dragicevic.

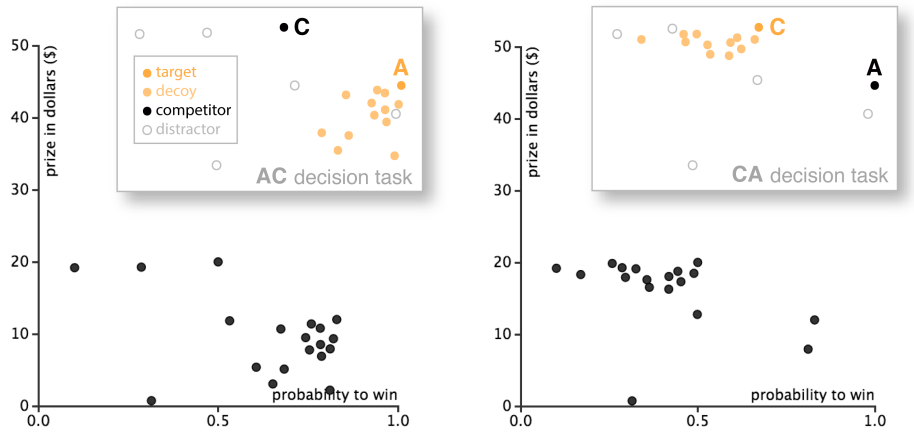


Figure 15. Illustration of the extension of the attraction effect in large datasets. Example of two matched decision tasks AC and CA in scatterplots.

The attraction effect is a well-studied cognitive bias in decision making research, where one's choice between two alternatives is influenced by the presence of an irrelevant (dominated) third alternative. We examine whether this cognitive bias, so far only tested with three alternatives and simple presentation formats such as numerical tables, text and pictures, also appears in visualizations. Since visualizations can be used to support decision making — e.g., when choosing a house to buy or an employee to hire — a systematic bias could have important implications. In a first crowdsource experiment, we indeed partially replicated the attraction effect with three alternatives presented as a numerical table, and observed similar effects when they were presented as a scatterplot. In a second experiment, we investigated if the effect extends to larger sets of alternatives, where the number of alternatives is too large for numerical tables to be practical. Our findings indicate that the bias persists for larger sets of alternatives presented as scatterplots. We discuss implications for future research on how to further study and possibly alleviate the attraction effect.

More on the project Web page: www.aviz.fr/decoy.

7.11. An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents

Participants: Pascal Goffin [correspondant], Jeremy Boy, Wesley Willett, Petra Isenberg.

We investigated the design and function of word-scale graphics and visualizations embedded in text documents. Word-scale graphics include both data-driven representations such as word-scale visualizations and sparklines, and non-data-driven visual marks. There has been little research attention on their design, function, and use so far. We present the results of an open ended exploratory study with nine graphic designers. The study resulted in a rich collection of different types of graphics, data provenance, and relationships between text, graphics, and data. Based on this corpus, we present a systematic overview of word-scale graphic designs, and examine how designers used them. We also discuss the designers' goals in creating their graphics, and characterize how they used word-scale graphics to visualize data, add emphasis, and create alternative narratives. We discuss implications for the design of authoring tools for word-scale graphics and visualizations building on these examples, and explore how new authoring environments could make it easier for designers to integrate them into documents.

Figure 16. Examples of word-scale graphics collected during the study.

CEDAR Team

7. New Results

7.1. Scalable Heterogeneous Stores

To improve data querying performance within polystores (Section 3.1), we developed Estocada, a novel system capable of exploiting side-by-side a practically unbound variety of data management system, all the while guaranteeing the soundness and completeness of the store, and striving to extract the best performance out of the various DMSs. Estocada leverages recent advances in the area of query rewriting under constraints, which we use to capture the various data models and describe the fragments stored within each data management system. Estocada was demonstrated at the IEEE ICDE conference [12]; recent experimental results demonstrated performance improvements by many orders of magnitude brought by the fragments Estocada supports, with respect to the setting where data is stored only in the system it originates from. This work continues, in collaboration with Alin Deutsch and Rana Alotaibi from UCSD.

7.2. Semantic Query Answering

This is a core topic for the team, in which the year has been particularly fruitful.

First, we investigated efficient query answering techniques in knowledge bases. A large and useful set of ontologies enjoys FOL (first-order logic) reducibility of query answering, that is: answering a query q can be reduced to evaluating a certain first-order logic (FOL) formula (obtained from the query and ontology) against only the explicit facts. We devised a novel query optimization framework for ontology-based data access settings enjoying FOL reducibility. Our framework is based on searching within a set of alternative equivalent FOL queries, that is, FOL reformulations, one with minimal evaluation cost when evaluated through a relational database system. We applied this framework to the DL-Lite$_{\mathcal{R}}$ Description Logic underpinning the W3C's OWL2 QL ontology language, and demonstrated through experiments its performance benefits when two leading SQL systems, one open-source and one commercial, are used for evaluating the FOL query reformulations. This work has lead to a major publication in the PVLDB journal [13], and a demonstration at the Semantic Web conference [4], while the complete details appear in [16] and the PhD thesis of the student author. [2].

Second, we initiated a study of extensions of conjunctive queries to conjunctive regular path queries. The first step has been to study regular path queries under linear existential rules, generalizing previous work on DL-Lite$_{\mathcal{R}}$, which is at the core of the Semantic Web OWL 2 QL profile. Regular path queries are queries that check for a path between two individuals, which is labeled by a word belonging to a given regular language. Such navigational languages are very popular for graph-based data representation, such as RDF. We have studied the complexity for this query language, and shown that it is NL-complete in data complexity, and EXPTIME-complete in combined complexity (and PTIME complete with bounded arity). This work has received the best paper award at RR'16 , and is currently being extented to conjunctive regular path queries.

Last, we studied the expressivity of several variants of Datalog, the classical language for deductive databases. In particular, we have studied its expressivity when given access (or not) to input negation (the ability to check if an extensional atoms hold or not) and to a linear order. We provided a complete Venn diagram regarding the expressivity of all the variants when considering homomorphism-closed query. The trickiest (and most surprising) points is the existence of polynomial-time computable homomorphism closed queries that are not expressible within Datalog with linear order but without input negation. These results have been published at IJCAI'16 [7].

7.3. Multi-model Querying

We have proposed a lightweight data integration architecture implemented within Tatooine (see Section 6.1.5); the system was demonstrated on a data journalism use case at the prestigious VLDB conference [9].

A separate effort in the area of multi-model querying considered querying databases of interconnected documents, users and concepts, by means of keywords. In this context, it is important that query results reflect not only the keywords present in documents but also the links between users and documents (so as to return to one user first the results authored in his social neighborhood), links between documents (for instance when a tweet answers another or an article has a link to another), and last but not least semantic information which allows interconnecting and interpreting terms mentioned in text. This research was finalized as part of the PhD of Raphaël Bonaque [1] and appeared at the EDBT conference 2016 [11].

7.4. Interactive Data Exploration at Scale

In the work with Enhui Huang (PhD student at Ecole Polytechnique), we seek to minimize the number of samples presented to the user for reviewing in order to build an accurate model of the user interest. In particular, as the dimensionality of the data space increases, the number of samples needed to build an accurate user interest model increases fast. We examine a range of popular feature selection techniques for data exploration, and for the best-performing feature selection technique, Gradient boosting regression trees (GBRT), we propose optimizations to overcome the issue of unbalanced training data and to dynamically determine the number of relevant features to select. Experimental results show that our optimized GBRT improves F-measure from nearly 0 without feature selection, to high F-measure (>0.8), by adaptively choosing the number of relevant features.

This work is currently under submission to a database conference.

7.5. Exploratory Querying of Semantic Graphs

We have started work with an intern (Zheng Zhang) toward automatically exploring the structure of an RDF graph and visualizing it with the help of a D3.js (https://d3js.org/) visualization library. These initial steps should serve to guide the beginning of an interactive exploration of the RDF graph in order to identify interesting analytical queries to be asked and evaluated. This work continues.

Separately, with a different intern (Javier Letelier), we have investigated efficient algorithms for keyword search in an RDF graph, exploiting structural and semantic knowledge about the graph; such knowledge is organized as an RDF summary which is an RDF graph itself. The algorithm was implemented and integrated as a text search tool within the Tatooine prototype; the work is ongoing.

DAHU Project-Team

6. New Results

6.1. Specification and verification of data-driven systems

Verification of Hierarchical Artifact Systems

Data-driven workflows, of which "business artifacts" are a prime exponent, have been successfully deployed in practice, adopted in industrial standards, and have spawned a rich body of research in academia, focused primarily on static analysis. Over the past few years, we have embarked upon a study of the verification problem for artifact systems. This is a challenging problem because of the presence of unbounded data. In order to deal with the resulting infinite-state system, we developed in earlier work a symbolic approach allowing a reduction to finite-state model checking and yielding a pspace verification algorithm for the simplest variant of the model (no database dependencies and uninterpreted data domain). Subsequently, we extended our approach to allow for database dependencies and numeric data testable by arithmetic constraints. In [19], we make significant progress on several fronts, by considering a much richer and more realistic model than in previous work, incorporating core elements of IBM's successful Guard-Stage-Milestone model. In particular, the model features task hierarchy, concurrency, and richer artifact data. It also allows database key and foreign key dependencies, as well as arithmetic constraints. The results require qualitatively novel techniques, because the reduction to finite-state model checking used in previous work is no longer possible. Instead, the richer model requires the use of a hierarchy of Vector Addition Systems with States. The arithmetic constraints are handled using quantifier elimination techniques, adapted to our setting.

Process-centric views of data-driven workflows.

We also studied the models of *data Petri nets* and *ν-Petri nets*. While these models were introduced in the verification community to analyse protocols and process algebra, they can also be seen as (very limited) data-driven workflows with only unary predicates. Our results this year show that various boundedness problems (e.g. can the database grow unbounded?) are decidable in data Petri nets [22], and pinpoint the exact complexity of safety analysis in ν-Petri nets [23].

Complexity in counter systems and in proof systems.

The static analysis of queries on XML trees and data streams relies in a majority of cases on decision procedures expressed in terms of formal systems like counter systems or proof systems. For instance, two-variables first-order data queries on words can be related to reachability in vector addition systems (VAS), and the same queries on trees to reachability in a branching extension of VAS [12]. We are at the forefront on the complexity analysis for such systems [15], [13], [16], [14].

We investigate in the ANR PRODAQ project a different angle on the static analysis of queries, relying on proof systems. Our first results on the subject [18] provide a sequent calculus for a modal data logic with an optimal proof-search algorithm.

6.2. Personal information management.

Thymeflow We developed Thymeflow, a personal knowledge base with spatio-temporal data [24].

The typical Internet user has data spread over several devices and across several online systems. We demonstrate an open-source system for integrating user's data from different sources into a single Knowledge Base. Our system integrates data of different kinds into a coherent whole, starting with email messages, calendar, contacts, and location history. It is able to detect event periods in the user's location data and align them with calendar events. We will demonstrate how to query the system within and across different dimensions, and perform analytics over emails, events, and locations.

EX-SITU Team

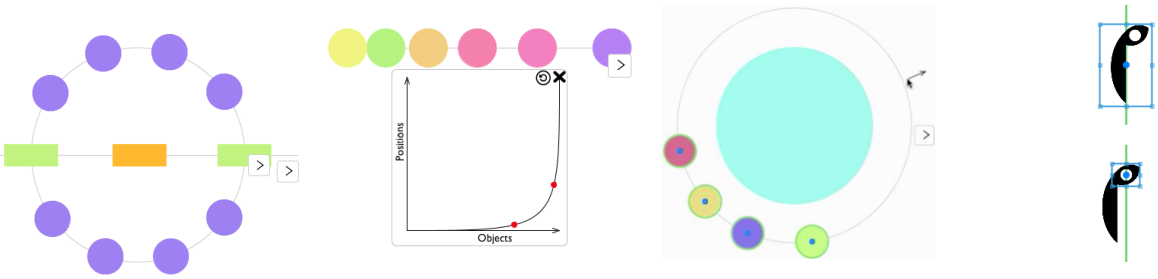
7. New Results

7.1. Fundamentals of Interaction

Participants: Sarah Fdili Alaoui, Michel Beaudouin-Lafon, Cédric Fleury, Wendy Mackay, Theophanis Tsandilas.

In order to better understand fundamental aspects of interaction, ExSitu studies interaction in extreme situations. We conduct 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.

StickyLines – Aligning and distributing graphical objects is a common, but cumbersome task. We studied graphic designers and regular users and identified three key problems with current tools: lack of persistence, unpredictability of the results, and inability to 'tweak' the layout. We created *StickyLines* [14], a tool that reifies guidelines into first-class objects: Users can create precise, predictable and persistent interactive alignment and distribution relationships, and can 'tweak' the alignment in a way that can be maintained for subsequent interactions (Figure 2). We ran a [2x2] within-participant experiment to compare *StickyLines* with standard commands and found that *StickyLines* performed up to 40% faster and required up to 50% fewer actions than traditional alignment and distribution commands for complex layouts. Finally, we gave *StickyLines* to six professional designers and found that not only did they quickly adopt it, they also identified novel uses, including creating complex compound guidelines and using them for both spatial and semantic grouping. This work demonstrate the power of reifying concepts, such as alignment and distribution, into first-class objects that can be directly manipulated and appropriated by end users.



(a) (b) (c) (d)

Figure 2. StickyLines reify alignment and distribution into first-class graphical objects that users can manipulate directly. (a) Circular and horizontal alignments. (b) Non-linear distribution. (c) Ghost guideline. (d) Tweaking an object's bounding box. .

UIST Video Browser – We created an interactive video browser that provides a rapid overview of the 30-second video previews of the ACM UIST conference papers, based on the conference schedule [16]. The web application was made available to the 600+ conference attendees, who could see an overview of upcoming talks, search by topic, and create personalized, shareable video playlists that capture the most interesting or relevant papers. Reifying playlists into first-class objects and applying instrumental interaction concepts helped create a fluid and efficient interface.

In(SITE) – We explored touch-based 3D interaction in the situation where users are immersed in a 3D virtual environment and move in front of a large multi-touch wall-sized display. We designed In(SITE) [20], a bimanual touch-based technique combined with object teleportation features which enables users to perform 3D object manipulation on a large vertical display (Figure 3). This technique was compared with a standard 3D interaction technique. The results showed that participants can reach the same level of performance for completion time and a better precision for fine adjustments with the In(SITE) technique. They also revealed that combining object teleportation with both techniques improves translation tasks in terms of ease of use, fatigue, and user preference.

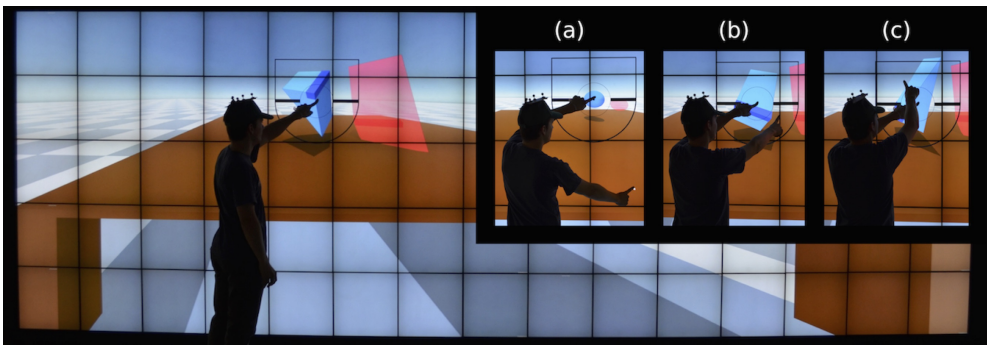


Figure 3. 3D manipulation on a multi-touch wall-sized display combining bimanual interaction and teleportation. The user is performing a xy translation (main pict.), z translation (a), roll rotation (b), and pitch & yaw rotation(c).

In collaboration with Inria Lille (MJOLNER group) and Univ. Strasbourg, we applied our design principles for instrumental interaction to create new interactive tools for the parallelization of programs, a highly specialized task that is currently done by expert developers. Current programming models, languages and tools do not help developers restructure existing programs for more effective execution. At the same time, automatic approaches are overly conservative and imprecise to achieve sufficient performance. We introduced *interactive program restructuring* [28], [11] to bridge the gap between semi-automatic program manipulation and software visualization. First, we extended a state-of-the-art polyhedral model for program representation so that it supports high-level program manipulation. Based on this model, we designed and evaluated a direct manipulation visual interface for program restructuring. This interface provides information about the program that was not immediately accessible in the code and allows to manipulate programs without rewriting code. By providing a visual and textual representation of an automatically computed program optimization that is easily modifiable and reusable by the developer, we create a sort of human-machine partnership where the developer can better take advantage of the power of the machine. An empirical study of developers using this tool showed the value of program manipulation tools based on the instrumental interaction paradigm. This work illustrates how the combination of our conceptual approaches, namely instrumental interaction and human-computer partnership, can benefit extreme users such as developers of parallel programs.

Finally, we reviewed statistical methods for the analysis of user-elicited gestural vocabularies [24]. We showed that measures currently used to assess agreement between participants of a gesture elicitation study are problematic. We discussed the problem of chance agreement and showed how it can bias results. We reviewed chance-corrected agreement coefficients that are routinely used in inter-reliability studies and showed how to apply them to gesture elicitation studies. We also discussed how to compute interval estimates for these coefficients and how to use them for statistical inference.

7.2. Partnerships

Participants: Wendy Mackay, Jessalyn Alvina, Ghita Jalal, Joseph Malloch, Nolwenn Maudet.

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 algortithm, 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 sytstems' that are discoverable, appropriable and expressive for the user. *Interactive program restructuring* [28] offers a concrete example, where expert programmers interact with dynamic visualisations of parallel programs to better understand and organize their code. Similarly, tools such as *Color Partner* generate color suggestions based the users input, helping the user guide their discovery of new color possibilities, and *Linkify* helps users create rules to define how visual properties should change under different user contexts (see Jalal's dissertation).

We hosted the 30-person *ERC CREATIV* workshop in Paris, to explore our concepts of *Co-adaptive Systems* (including human-centered machine learning); and *Instrumental Interaction* (including substrates) with prominent researchers from Stanford University, New York University, University of Aarhus, Goldsmiths College, University of Toulouse, IRCAM, University of British Columbia, UC San Diego, and UC Berkeley. Our long-term, admittedly ambitious, goal is to create a unified theory of interaction grounded in how people interact with the world. Our principles of co-adaptive systems and instrumental interaction offer a generative approach for supporting creative activities, from early exploration to implementation. The workshop launched several research projects that are currently in progress or will be published in 2017.

Human-Centred Machine Learning:

We begin by challenging some of the standard assumptions surrounding Machine Learning, clearly one of the most important and successful techniques in contemporary computer science. It involves the statistical inference of models (such as classifiers) from data. However, all too often, the focus is on impersonal algorithms that work autonomously on passively collected data, rather than on dynamic algorithms that progressively reveal their progress to support human users. We collaborated on a workshop at the CHI 2017 conference, entitled "Human-centred Machine Learning" [15] with colleagues from Ircam, Goldsmiths College, and Microsoft Research. We seek a different understanding of the 'human-in-the-loop', where the focus is less on the human user as input to an algorithm, but rather as an algorithm in service of a human user. Examining machine learning from a human-centred perspective includes explicitly recognising human work in the creation of these algorithms, as well as the situated use these algorithms by human work practices. A human-centred understanding of machine learning in human context can lead not only to more usable machine learning tools, but to new ways of framing learning computationally.

Supporting Expressivity:

We helped organize and participated in a workshop at CHI 2017 *Human Computer Interaction meets Computer Music* [27], where we described the results of the MIDWAY Equipe Associé project (with McGill University, Ex-Situ and the *MINT* EP at Inria, Lille.) We presented results of our extensive research with contemporary music composers, in particular our strategy for developing 'co-adaptive instruments'. This involves a paradigm shift, where the goal of the technology is not necessarily the accuracy of a particular result, but rather, the human user's ability to express themselves through the technology.

We also explored the idea of rethinking the use of machine learning to support human-computer partnerships for everyday interaction. We built on gesture-typing, which offers users an efficient, easy-to-learn, and error-tolerant technique for producing typed text on a soft keyboard. Our focus was not on improving recognition accuracy, which we take as a given, but rather on how to make gesture-typed output more expressive. Experiment 1 demonstrated that users vary word gestures according to instructions (accurately, quickly or creatively) as well as specific characteristics of each word, including length, angle, and letter repetition. We show that users produce highly divergent gestures, with three easily detectable characteristics: curviness, size, and speed. We created the Expressive Keyboard [10] which maps these characteristics to color variations, thus allowing users to control both the content and the color of gesture-typed words (Figure 4). Experiment 2 demonstrates that users can successfully control their gestures to produce the desired colored output, and find it easier to react to visual feedback than explicitly controlling the characteristics of each gesture. Expressive

keyboards can map gestural input to any of a variety of output characteristics, such as personalized handwriting and dynamic emoticons, to let users transform gesture variation into expressivity, without sacrificing accuracy.

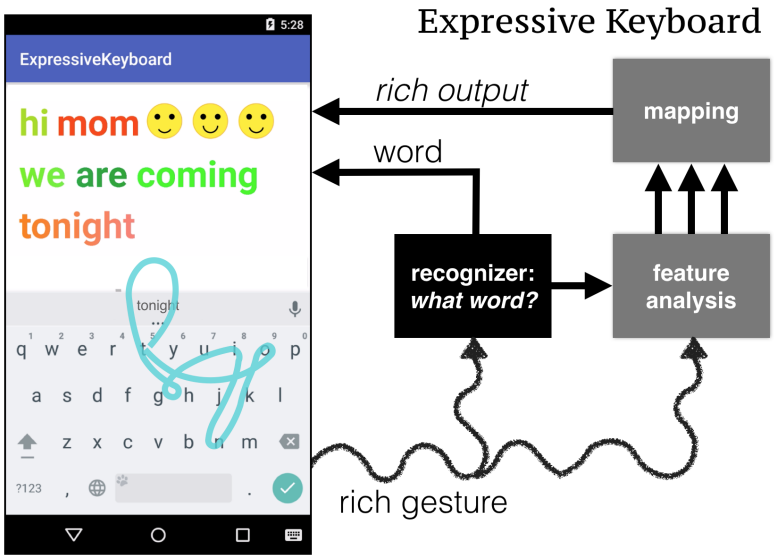


Figure 4. Expressive Keyboards produce accurate words, but also let users control multiple expressive output properties.

7.3. Creativity

Participants: Sarah Fdili Alaoui, Michel Beaudouin-Lafon, Ghita Jalal, Wendy Mackay, Joseph Malloch, Nolwenn Maudet, Michael Wessely, 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 explore how concepts of substrate and co-adaptation can change how we design interactive technology for supporting creativity. Co-adaptation is the phenomenon in which users both adapt their behavior to the system's constraints, and appropriate the system for their own needs. We explore these concepts using participatory design studies in creative contexts with expert and non-expert users. We study structuring layouts for graphic designers, sketching movement for choreographers, expressive movements for dancers and further explore expressive gesture of non-experts on mobile devices and possible interactions on hybrid stretchable interfaces. These studies require a multi-disciplinary design team that works closely with users throughout the design process. We create situations that cause users to reflect deeply about their activities in context and work with them to articulate the design problem. The experiments, prototypes and systems that we developed and deployed are illustrated below:

Graphic design: Our studies of the creative design practices of professional graphic designers show that designers appropriate visual properties of existing tools to create their own personal 'instruments'. Unfortunately, most professional design tools make this difficult: At best, they provide only indirect access, through property sheets or dialog boxes, to visual properties, such as color and style, rather treating them as as independent interactive objects. We developed a number of composition tools that demonstrate how to explicitly reify visual properties, using the concept of co-adaptive instruments. Ghita Jalal successfully defended her doctoral dissertation on this topic (see [9]).

We also examined artists' and designers' practices as they manipulate color and create layouts in their projects. We found that artists and designers select colors from personal representations. They manipulate color in the context of its surrounding graphical elements, and combine it with other visual properties such as texture. As they create their layouts, designers establish links among visual properties such as size, position, and layering of graphical elements. They define rules for how these properties change in space, across instances of the same composition, or in time, across related compositions. We also found that designers prefer tools that provide direct access to visual properties.

Choreography: We are interested in designing choreographic support tools because choreographers rarely have access to interactive tools that are designed specifically to support their creative process [13]. In order to design for such a technology, we interviewed six contemporary choreographers about their creative practice. We found that even though each process is unique, choreographers represent their ideas by applying a set of operations onto choreographic objects. Throughout different creative phases, choreographers compose by shifting among various degrees of specificity and vary their focal points from dancers to stage, to interaction, to the whole piece. Based on our findings, we presented a framework for articulating the higher-level patterns that emerge from these complex and idiosyncratic processes. We then articulated the resulting implications for the design of interactive tools to support the choreographic practice.

On generating choreographic ideas, we developed the Choreographer's Workbench, a full-body interactive system that aims to help choreographers explore and design dance movements during the ideation phase by creating a link between past recorded movement ideas and revealing their underlying relationships. The system explores how to increase the discoverability and appropriateness of movement ideas via feedforward visualization of movement characteristics.

Figure 5. Passersby interact with the animated Père Noël, first mirrors their behavior and then shapes it.

We collaborated with the N+1 theater group on the "Grande Vitrine" art and science project, an interactive installation that takes place during the month of Christmas. It consisted of a virtual animated character with whom participants interact and a physical kinetic sculpture whose motions are triggered by participant interaction (Figure 5). The participant were expected to perform full-body movements and figure out the correct one that will help the animated character escape from the virtual screen into the physical motorized

display. The installation tested the concept of ”shaping” from experimental psychology where the participant is guided to make ”successive approximations” in arriving at the correct gesture. It was installed at the theater of Évry, that has a display on the shopping mall in Évry for the entire month of December.

Finally, we collaborated with Simon Fraser University on an interactive installation called still, moving. The installation created a sonic experience that heightens self-awareness of our micro-movements in stillness. Sound created an intimate envelope that nurtures self-reflection and the experience of inward sensations. In still, moving, the audience was equipped with two Myo Armbands that capture their movements as well as their muscular activity. The physiological signals such as muscle tension and subtle accelerations were analyzed and mapped to a sound environment in order to increase perception of the inner self. The design of the relationship between movement and sound was evolving along the interaction, shifting the soundscape from reflective to challenging, guiding the audience in an exploration of novel and gradual relationship to weight and understanding of the complexity of the silent body.

Everyday creativity:

Finally, for non expert users, we developed an inexpensive method for fabricating *Stretchis*, highly stretchable interfaces that combine sensing and displaying capabilities [22]. This method enables designers and casual makers to embed transparent conductors and electroluminescence displays in stretchable PDMS substrates (Figure 6). We showed how to prototype stretchable user interfaces for a range of application scenarios by using standard design software and screen-printing techniques. Despite the use of inexpensive equipment, our results demonstrate that we can produce durable and highly stretchable sensors and displays that remain functional under strain levels of more than 100%.

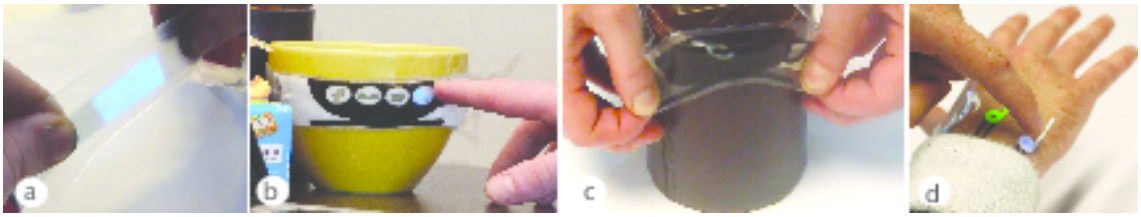


Figure 6. Stretchis are highly stretchable user interfaces that include touch and proximity sensors and electroluminescent displays (a). Stretchis are transparent (b); can be stretched to fit to the geometry of different physical objects (c); and can act as on-skin user interfaces (d).

7.4. Collaboration

Participants: Michel Beaudouin-Lafon, Cédric Fleury, Wendy Mackay, Can Liu, Ignacio Avellino Martinez.

ExSitu is interested in exploring new ways to support collaborative interaction, especially within and across large interactive spaces such as those of the Digiscope network (http://digiscope.fr/).

We studied how wall-sized displays support small groups of users working together on large amounts of data. We conducted observational studies showing that users adopt a range of collaboration styles, from loosely to closely coupled and that shared interaction techniques, in which multiple users perform a command collaboratively, support co-located collaborative work. In order to test the effect of such shared interaction techniques, we operationalize five collaborative situations with increasing levels of coupling in a data manipulation task [18]. The results show the benefits of shared interaction for close collaboration: it encourages collaborative manipulation, it is more efficient and preferred by users, and it reduces physical navigation and fatigue. We also identified the time costs caused by disruption and communication in loose collaboration and analyzed the trade-offs between parallelization and close collaboration. Altogether, these findings can inform the design of shared interaction techniques to support collaboration on wall-sized displays.

We are also interested in how to help teams of novice crafters prototype physical objects. To this end, we conducted a study [12] framed around two all-day design charrettes where novices performed a complete design process: ideation sketching, concept development and presentation, fabrication planning documentation and collaborative fabrication of hand-crafted prototypes. This structure allowed us to control key aspects of the design process while collecting rich data about creative tasks, including sketches on paper, physical models, and videos of collaboration discussions. Participants used a variety of drawing techniques to convey 3D concepts. They also extensively manipulated physical materials, such as paper, foam, and cardboard, both to support concept exploration and communication with design partners. Based on these observations, we proposed design guidelines for CAD tools targeted at novice crafters.

ILDA Project-Team

7. New Results

7.1. Wall Displays

Ultra-high-resolution wall displays feature a very high pixel density over a large physical surface, which makes them well-suited to the collaborative, exploratory visualization of large datasets (see Sections 6.3.1 and 6.3.2). We have continued working on the design, implementation and evaluation of interactive visualization techniques for such ultra-high-resolution wall-sized displays, focusing, in some of these projects, on the collaboration between users who perform different data manipulation and analysis tasks.

Figure 4. ***Left:*** *FITS-OW running on the WILDER platform, showing: multiple FITS images, (a) M31 on the left side, (b) three juxtaposed images that show observations of the Eagle nebula at different wavelengths, and (c) a much larger FITS image (86,499 × 13,474 pixels) used as a zoomable background over the entire wall; (d) the result-set of a SIMBAD query restricted to observations about galaxies; (e) basic measurements for galaxy M31; (e) a page of a research paper (PDF) discussing that particular galaxy; (f) the color map selector.* ***Right:*** *Results of a SIMBAD query superimposed on the corresponding FITS image, along with a sorted list of all items in the result-set. Selecting an element in this list updates the detailed info in the lower right window and highlights the source in the image. All windows can be freely repositioned on the wall.*

- We continued working on FITS-OW, an application designed for such wall displays, that enables astronomers to navigate in large collections of FITS images, query astronomical databases, and display detailed, complementary data and documents about multiple sources simultaneously. We published a paper about FITS-OW [7], in which we describe the system, reporting on the technical challenges we addressed in terms of distributed graphics rendering and data sharing over the computer clusters that drive wall displays. The article also describes how astronomers interact with their data using both the wall's touch-sensitive surface and handheld devices. This work was also featured as a short article in the SPIE Newsroom (see Section 10.3).
- Wall-sized displays support small groups of users working together on large amounts of data. Observational studies of such settings have shown that users adopt a range of collaboration styles, from loosely to closely coupled. Shared interaction techniques, in which multiple users perform a command collaboratively, have also been introduced to support co-located collaborative work. In [19], we operationalized five collaborative situations with increasing levels of coupling, and tested the effects of providing shared interaction support for a data manipulation task in each situation. The results show the benefits of shared interaction for close collaboration: it encourages

collaborative manipulation, it is more efficient and preferred by users, and it reduces physical navigation and fatigue. We also identify the time costs caused by disruption and communication in loose collaboration and analyze the trade-offs between parallelization and close collaboration. These findings inform the design of shared interaction techniques to support collaboration on wall-sized displays.

- We also studied how pairs explore graphs on a touch enabled wall-display [16], using two selection techniques adapted for collaboration: a basic localized selection, and a propagation selection technique that uses the idea of diffusion/transmission from an origin node. We assessed in a controlled experiment the impact of selection technique on a shortest path identification task. Pairs consistently divided space even though the task is not spatially divisible. The basic selection technique, that has a localized visual effect, led to parallel work that negatively impacted accuracy. The large visual footprint of the propagation technique led to close coordination, improving speed and accuracy for complex graphs only. We then observed the use of propagation on additional graph topology tasks, confirming pair strategies on spatial division and coordination.
- In [22], we focused on road traffic control center. Road traffic control centers are of vital importance to modern cities. Interviews with controllers in two such centers identified the need to incorporate the visualization of results from predictive traffic models with real traffic, to help operators choose among different interventions on the network. We explore this idea in a prototype that runs on a wall display, and supports direct touch and input from workstations and mobile devices. Apart from basic functionality to manage the current traffic such as changing traffic light duration or speed limits, the prototype incorporates traffic simulations for forecasting results of possible actions, highlighting their differences to current traffic. Based on needs identified in our interviews, we offered two techniques that visually combine simulated and real situations, taking advantage of the large display space: multiple independent views and DragMagic, a variation of magic lenses. A preliminary laboratory experiment suggests that both techniques are viable design options, even for monitoring several simulations and areas of interest, contrary to expectations from previous work. However DragMagics are easier to master.
- Immersion is the subjective impression of being deeply involved in a specific situation, and can be sensory or cognitive. In a position paper [23], we used a basic model of visual perception to study how ultra-high resolution wall displays can provide visual immersion. With their large size, depending on the position of viewers in front of them, wall displays can provide a surrounding and vivid environment. Users close to the wall can have their visual field filled by the wall and they are able to see clearly a large amount information with a fine resolution. However, when close to the wall, visual distortion due to large possible viewing angles, can affect the viewing of data. On the contrary, from far away, distortion is no longer an issue, but the viewers' visual field is not fully contained inside the wall, and the information details seen are less fine.

7.2. Gestures, Tangibles and Sound

- We designed a new way of implementing tangible interfaces with TouchTokens [4]. The approach requires only 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. We performed a formative user study to collect and analyze touch patterns with tokens of varying shape and size. The analysis of this pattern collection showed that individual users have a consistent grasp for each token, but that this grasp is user-dependent and that different grasp strategies can lead to confounding patterns. We thus designed a second set of tokens featuring notches that constrain users' grasp. Our recognition algorithm can classify the resulting patterns with a high level of accuracy ($>95\%$) without any training, enabling application designers to associate rich touch input vocabularies with command triggers and parameter controls.

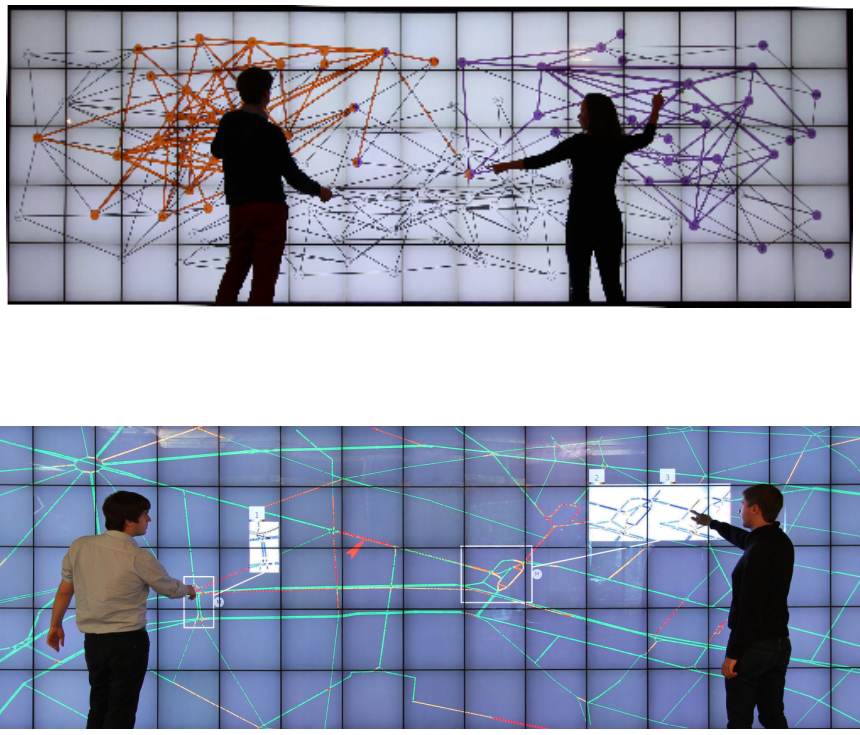


Figure 5. ***Top:*** *A pair using the propagation technique described in [16] to explore a graph. They discuss two communities, in orange and purple, selected using the propagation technique. The communities are linked by a specific node shown by the right user. The remaining 3 orange-purple nodes show how by propagating the purple community, it flows into the orange one through this node.* ***Bottom:*** *Visualization from [22] of traffic in a city with two "DragMagics" (white rectangles) showing one (left) and two (right) simulations associated with different possible interventions on the traffic. The simulation visualizations use difference color maps to highlight differences with the real traffic.*

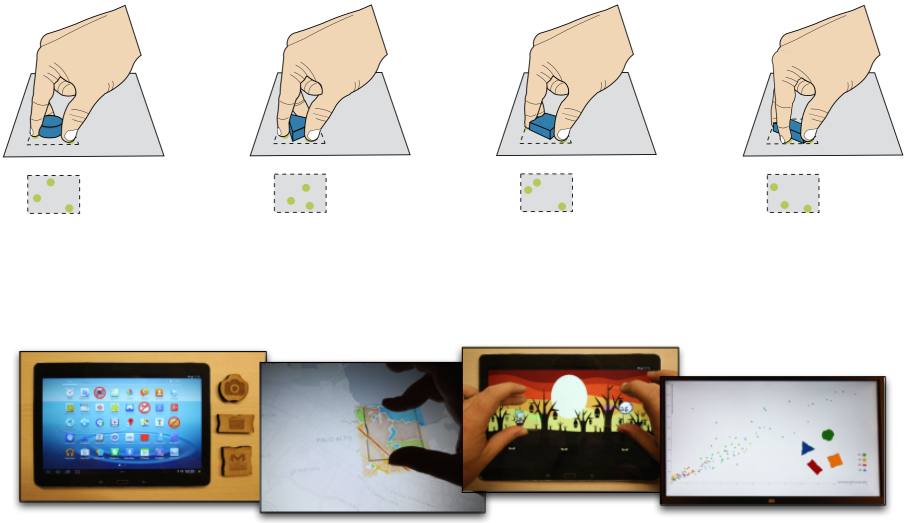


Figure 6. ***Top:*** *TouchTokens are passive tokens that guide users' fingers to specific spatial configurations, resulting in distinguishable touch patterns.* ***Bottom:*** *Proof-of-concept applications: access control, tangible magic lenses, character controllers in a game, data visualization.*

- In collaboration with IRCAM, we introduced SoundGuides [17], a user adaptable tool for auditory feedback on movement. The system is based on a interactive machine learning approach, where both gestures and sounds are first conjointly designed and conjointly learned by the system. The system can then automatically adapt the auditory feedback to any new user, taking into account the particular way each user performs a given gesture. SoundGuides is suitable for the design of continuous auditory feedback aimed at guiding users' movements and helping them to perform a specific movement consistently over time. Applications span from movement-based interaction techniques to auditory-guided rehabilitation. We first describe our system and report a study that demonstrates a "stabilizing effect" of our adaptive auditory feedback method.

7.3. Interacting with Linked Data

As part of the team's novel research theme on Semantics-Driven Data Manipulation 3.2 , and in collaboration with Aba-Sah Dadzie from the Open University, Emmanuel Pietriga coordinated a special issue of the Semantic Web Journal and wrote a follow-up [12] to the 2011 survey about Approaches to Visualizing Linked Data [42]. Linked Data promises to serve as a disruptor of traditional approaches to data management and use, promoting the push from the traditional Web of documents to a Web of data. The ability for data consumers to adopt a follow your nose approach, traversing links defined within a dataset or across independently-curated datasets, is an essential feature of this new Web of Data, enabling richer knowledge retrieval thanks to synthesis across multiple sources of, and views on, interrelated datasets. But for the Web of Data to be successful, we must design novel ways of interacting with the corresponding very large amounts of complex, interlinked, multi-dimensional data throughout its management cycle. The design of user interfaces for Linked Data, and more specifically interfaces that represent the data visually, play a central role in this respect. Contributions to this special issue on Linked Data visualization investigate different approaches to harnessing visualization as a tool for exploratory discovery and basic-to-advanced analysis. The papers in this volume illustrate the design and construction of intuitive means for end-users to obtain new insight and gather more knowledge, as they follow links defined across datasets over the Web of Data.

7.4. Visualization

- The attraction effect is a well-studied cognitive bias in decision making research, where one's choice between two alternatives is influenced by the presence of an irrelevant (dominated) third alternative. In collaboration with EPI Aviz, we examined in [13] whether this cognitive bias, so far only tested with three alternatives and simple presentation formats such as numerical tables, text and pictures, also appears in visualizations. In a series of crowdsource experiments, we observed this cognitive bias in visualizations (namely scatterplots), even in larger sets of alternatives, never considered before, where the number of alternatives is too large for numerical tables to be practical. We discussed implications for future research on how to further study and possibly alleviate the attraction effect.
- With colleagues from the University of Konstanz [14] we concluded previous work on data glyphs, i.e., visual marks that encode multiple dimensions to one or more visual variables. We provided a systematic review of experimental studies on data glyphs from the past 60 years, describing the types of glyphs and design variations tested, the tasks under which they were analyzed, and study results. Based on our meta analysis of all results, we further contributed a set of design implications and a discussion on open research directions.
- In [11], with colleagues from INRA, we provided an overview of a framework for Evolutionary Visual Exploration (EVE) that guides users in exploring large search spaces. EVE uses an interactive evolutionary algorithm to steer the exploration of multidimensional datasets towards two-dimensional projections that are interesting to the analyst. Our method smoothly combines automatically calculated metrics and user input in order to propose pertinent views to the user. While previously we showed that using EVE, domain experts were able to formulate interesting hypothesis and reach new insights when exploring freely, our new findings indicate that users, guided by the interactive evolutionary algorithm, are able to converge quickly to an interesting view of their data when a clear task is specified. Our work aims at building a bridge between the domains of visual analytics and interactive evolution.

SMIS Project-Team

6. New Results

6.1. Embedded Data Management

Participants: Nicolas Anciaux, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa [correspondent].

Embedded keyword indexing: In this work, we revisit the traditional problem of information retrieval queries over large collections of files in an embedded context. A file can be any form of document, picture or data stream, associated with a set of terms. A query can be any form of keyword search using a ranking function (e.g., TF-IDF) identifying the top-k most relevant files. The proposed search engine can be used in sensors to search for relevant objects in their surroundings, in cameras to search pictures by using tags, in personal smart dongles to secure the querying of documents and files hosted in an untrusted Cloud, or in a personal cloud securely managed using a tamper resistant smart object. A search engine is usually based on a (large) inverted index and queries are traditionally evaluated by allocating one container in RAM per document to aggregate its score, making the RAM consumption linear with the size of the document corpus. To tackle this issue, we designed a new form of inverted index which can be accessed in a pure pipeline manner to evaluate search queries without materializing any intermediate result. Successive index partitions are written once in Flash and maintained in the background by timely triggering merge operations while files are inserted or deleted from the index. This work was initially published at VLDB'15 [5] and demonstrated at SIGMOD'15 [38]. It constitutes the main contribution of the PhD thesis of Saliha Lallali defended in January 2016. In 2016, we extended this work to demonstrate at EDBT'16 [22] its applicability to set up a secure distributed search engine for the Personal Cloud. We also complemented this work with (1) a thorough analysis of the RAM consumption linked to the main algorithms implementing the solution, (2) the support of conditional top-k queries in a personal Cloud context that we consider as a killer application domain today and (3) new performance measurements with a real dataset (ENRON), representative of this personal Cloud context. These new contributions have been submitted to Information Systems journal.

6.2. Secure Global Computing on Asymmetric Architecture

Participants: Benjamin Nguyen [correspondent], Axel Michel, Philippe Pucheral, Iulian Sandu Popa.

Asymmetric Architecture Computing: This research direction studies the secure execution of various algorithms on data stored in an unstructured network of Trusted Cells (i.e., personal trusted device) so that each user can keep control over her data. The data could be stored locally in a trusted cell or encrypted on some external cloud. Execution takes place on a specific infrastructure called the Asymmetric Architecture (AA): the network of trusted cells, supported by an untrusted cloud supporting IaaS or PaaS. Our objective is to show that many different algorithms and computing paradigms can be executed on AA, thus achieving secure and private computation. Our first contribution in this area was to study the execution of Privacy Preserving Data Publishing algorithms on such an architecture (T. Allard's PhD Thesis). Then we studied general SQL queries in this same execution context. We concentrated on the subset of SQL queries without joins, but including Group By and aggregates, and show how to secure their execution in the presence of honest-but-curious attackers. This work, named SQL-AA and notably published at EDBT'14 [8] and demonstrated at VLDB'15, was part of Quoc-Cuong To's Ph.D defended in 2015. We have extended this framework through a collaboration with INSA Centre Val de Loire, LIFO Lab and University of Paris Nord, LIPN lab and have shown in CoopIS'15 [9] that it is possible to achieve seamless integration of distributed MapReduce processing using trusted cells, while maintaining reasonable performance. In 2016, we added three novel contributions to SQL-AA: (i) an extended privacy analysis in which we consider stronger adversaries with more background knowledge, (ii) an extended threat model in which we consider malicious attacker and propose safety properties to prevent malicious attacks and (iii) we tackled practical issues like exchanging securely shared keys among trusted cells and Querier (GKE protocol) and enforcing access control at query

execution time. These new contributions have been published in TODS'16 [15]. In parallel, we are starting a new study in the line of our previous work on Privacy Preserving Data Publishing (PPDP) with the objective to inject individualized privacy requirements in the PPDP protocol. A preliminary contribution has been published at BDA'16 [25] to compute SQL aggregate queries under k-anonymity constraints where each individual contributing to the query may define her own k constraint, thereby letting each one weighting differently the sensitiveness of a given piece of information according to her own situation.

Secure spatio-temporal distributed processing: Mobile participatory sensing could be used in many applications such as vehicular traffic monitoring, pollution tracking, or even health surveying (e.g., to allow measuring in real-time the individual exposure to environmental risk factors or the propagation of an epidemic). However, its success depends on finding a solution for querying a large number of users which protects user location privacy and works in real-time [10]. We addressed these issues and proposed PAMPAS, a privacy-aware mobile distributed system for efficient data aggregation in mobile participatory sensing. In PAMPAS, mobile devices enhanced with secure hardware, called secure probes, perform distributed query processing, while preventing users from accessing other users' data. Secure probes exchange data in encrypted form with help from an untrusted supporting server infrastructure. PAMPAS uses two efficient, parallel, and privacy-aware protocols for location-based aggregation and adaptive spatial partitioning of secure probes. Our experimental results and security analysis demonstrate that these protocols are able to collect, aggregate and share statistics or derived data in real-time, without any privacy leakage. This work is part of Dai Hai Ton That's Ph.D. thesis defended in January 2016, co-supervised by Iulian Sandu Popa. The system implementation was demonstrated in [41], while two papers describing the technical details of the system have been published in 2016 [23], [16].

6.3. Personal Cloud

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Julien Loudet, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa, Guillaume Scerri, Paul Tran Van.

We are witnessing an exponential increase in the acquisition of personal data about the individuals or produced by them. Today, this information is managed using Web applications, centralizing this data in cloud data servers, under the control of few Web majors [2]. However, it has now become clear that (1) centralizing millions of personal records exposes the data to very sophisticated attacks, linked to a very high potential benefit in case of success (millions of records being revealed), and (2) delegating the management of personal records without any tangible guarantee for the individuals leads to privacy violations, the data being potentially made accessible to other organizations (e.g., governments, commercial partners) and being subject to lucrative secondary usages (not advertised to the individuals). To face this situation, many recent initiatives push towards the emergence of the Personal Cloud paradigm. A personal cloud can be viewed as a personal server, owned by a given individual, which gives to its owner the ability to store her complete digital environment, synchronize it among various devices and share it with other individuals and applications under control. In the SMIS team, we claim the need of a Secure Personal Cloud, and promote the introduction of a secure (tamper resistant) data engine in the architecture [30]. On this basis, we investigate new data sharing and dissemination models, where usage and access control rules endorsed by the individuals could be enforced and have presented this vision at EDBT'14 [6] and at ADBIS'15 [31]. We have started a cooperation with the startup CozyCloud at the end of 2014. A contract was signed at the end of 2014 to integrate PlugDB in a CozyCloud instance and two CIFRE PhD thesis have been launched so far. Paul Tran Van's PhD thesis explores a new data sharing paradigm dedicated to the personal cloud context. This paradigm, called SWYSWYK (Share What You See with Who You Know), allows to automatically derive intuitive sharing rules from a personal cloud content, to share rules among a community of users and to let each user physically visualize the net effects of these rules on her own Personal Cloud. We propose a reference architecture providing the users with tangible guarantees about the enforcement of SWYSWYK policies and demonstrate through a performance evaluation conducted on a real personal cloud platform that the approach is practical. This work constitutes the core of Paul Tran Van's thesis and is being submitted for publication at VLDB. Preliminary ideas related to this work are presented in ERCIM news'16 [27]. Julien Loudet's PhD thesis is just starting with the objective to explore privacy-preserving distributed computations over personal clouds.

More generally, the personal cloud context gain in importance in our research work. It is even at the heart of our future project-team named PETRUS (PErsonal and TRUSted cloud). PETRUS is expected to take over from the SMIS team beginning of 2017.

6.4. Interdisciplinary study on Privacy-by-Design

Participants: Nicolas Anciaux, Luc Bouganim [correspondent], Athanasia Katsouraki, Benjamin Nguyen, Philippe Pucheral.

The objective of this research action is to study the reciprocal entanglements between economic, legal, societal and technological aspects of the management and exploitation of personal data. Indeed, devising new ways of protecting data privacy cannot be done in isolation; it requires also identifying alternative economic models that are both viable and regulatory compliant. We started an interdisciplinary research work with economists (RITM Lab) and jurists (CERDI and DANTE labs) in the privacy axis of ISN (Institut de la Société Numérique) and plan to pursue it in two projects in preparation: the Convergence Institute I2DRIVE (Interdisciplinary Institute for Data Research: Intelligence, Values and Ethics) and the CNRS Federation SIHS (Sciences Informatiques, Humaines et Sociales) at UVSQ. A first interdisciplinary work conducted in 2016 concerns the design of a privacy preserving platform needed to conduct privacy studies “in vivo”. Such platforms are required to validate the effectiveness of privacy preserving solutions, in terms of technical feasibility, lawfulness, acceptability and benefits. To this end, we have designed a privacy preserving mobile lab, derived from the personal cloud platform developed by the team (see ‘Software’ section). In her PhD thesis, Athanasia Katsouraki developed a beta-version of that platform and used it to perform a pre-experimentation in the context of online form based survey, targeting 140 students. The goal was threefold: (1) to test the effectiveness of the proposed platform, (2) to test the adequation of the questionnaire and experimentation protocol (a result for the experimental economist), and (3) to check the impact of the use of a secure platform on the student’s answers. The pre-experimentation showed several improvement axis and led to the actual design of the privacy preserving mobile lab described in the Software section.

Another joint work is related to the design of technical means to help individuals perceive how their personal life is exposed compared to others and to make appropriate protection choices. This work led to the definition of a new principle called Privacy-by-Using [20], that we introduced to try to circumvent the limits of the privacy-by-design principle promoted by the regulator. The confrontation of the Privacy-by-Using principle with Big Data processing [26] has also been studied with jurists and economists.

Finally, we conducted a scientific expertise on behalf of DGCCRF (Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes) and of the European Council regarding the draft proposal of "Directive of the European Parliament and of the Council on certain aspects concerning contracts for the online and other distance sales of goods" regulating the payment of numeric goods and services by means of personal data. This led us to a cross-analysis, with researchers in Law and computer scientists, of technical, societal and economic issues linked to the smart disclosure principle, that is, under which conditions and formats individuals can get their data back from service providers [17], [19], [18].

6.5. Formal guarantees

Participant: Guillaume Scerri.

The aim of the action is to investigate the changes required for the PlugDB architecture to be amenable to formal security proofs.

More precisely we started exploring the precise formal guarantees that are desirable for a personal data server. Following work started in Bristol [7], relating to formal guarantees provided by secure hardware, we started studying how one could leverage the low level guarantees provided by secure hardware (PlugDB for example) to cover the more complex operations and guarantees required of a personal data server. The first finding of the action is that a modular architecture is required for formal proofs to be obtainable. This is reflected in the architectural concerns presented in the PETRUS project.

Additionally, we started studying how to leverage secure hardware guarantees in order to perform secure computations on distributed data. A first result in this direction is presented in [33], and submitted to Financial Cryptography 2017.