Activity Report 2015

Section New Results

<div align="center" style="color:red">ATEAMS Project-Team</div>

6. New Results

6.1. Faster Immutable Data Structures for the JVM

Immutable data structures involve copying when updating. Efficient implementations use persistent data-structures, so that most of the unchanged data is shared between the copies. Existing libraries for such data structures in the context of the Java virtual machine (JVM), such as the data structures in Clojure and Scala, are based on Hash Array-Mapped Tries (HAMTs), which provide efficient insertion and concatenation operations for persistent maps and sets. In [37] Steindorfer and Vinju presented additional optimisation which allow such operations to be up to 28 times faster than in the Clojure and Scala libraries. Furthermore, the cost of equality checking of such data structures is lower as well. All this, without incurring additional memory.

6.2. Automated Measurement and Analysis of Open Source Software

> Deciding whether an open source software (OSS) meets the required standards for adoption in terms of quality, maturity, activity of development and user support is not a straightforward process. It involves analysing various sources of information, including the project's source code repositories, communication channels, and bug tracking systems. OSSMETER extends state-of-the-art techniques in the field of automated analysis and measurement of open-source software (OSS), and develops a platform that supports decision makers in the process of discovering, comparing, assessing and monitoring the health, quality, impact and activity of opensource software. To achieve this, OSSMETER computes trustworthy quality indicators by performing advanced analysis and integration of information from diverse sources including the project metadata, source code repositories, communication channels and bug tracking systems of OSS projects [29], [26]

This result comes from intensive collaboration in the FP7 STREP project "OSSMETER". The ATEAMS contribution is focused around source code metrics and activity analysis for Java and PHP.

6.3. Modular Interpreters for the Masses

Object Algebras [46] are new design pattern for increased modularity and extensibility of tree based, abstract data types. By modelling the abstract syntax of a language as a generic factory interface, implementations of this interface provide multiple semantics of the data. For instance, one can define evaluation, type checking and pretty printing of the abstract syntax fully modularly. Additionally, the pattern allows syntax extension: adding a new constructor to the datatype, and modularly extending any existing interpretations to deal with the construct. The same interpretation of different constructs, however, might involve different kinds of context information. For instance, evaluation of arithmetic expressions does not require any context information, but evaluation of variables and binders requires and environment. In [34], Inostroza and Van der Storm introduce a simple, modular, and type safe technique to allow such interpretations to be composed anyway. It is based on lifting one interpreter to implicitly propagate the context information it does not require, so that the signatures of the interpreters become compatible. As a result, semantic definitions of language modules do not have to anticipate all kinds of context information that might be required by other modules with which it might be composed. The technique is simple, does not sacrifice separate compilation, is easy automate, and works in mainstream languages. It provides a first step towards a foundation for defining language by assembling modular building blocks.

6.4. One Parser to Rule Them All

Parsing realistic languages requires much more than just a parsing algorithm. Different kinds of language require advanced disambiguation, operator priorities, off-side rule checking, whitespace dependence or data dependence. In [25], Afroozeh and Izmaylova showed how most of these concerns are actually instances of data dependent parsing: the parsing process depends on the value of previously parsed input. They provided an encoding of indentation sensitive parsing, operator precedence and parsing in the presence of preprocessor directives, to a simple, data dependent core language which is executed using the general parsing algorithm GLL. By exposing the data dependent machinery at the level of the grammar formalism , this opens up a range of possibilities for custom parsing aspects, and provides a clear semantics for existing concerns like disambiguation.

6.5. A Pattern-Based Game Mechanics Design Assistant

Video game designers iteratively improve player experience by play testing game software and adjusting its design. Deciding how to improve gameplay, however, is difficult and time-consuming: designers lack an effective means for exploring decision alternatives and modifying a game's mechanics. In [35], Van Rozen presented the Mechanics Pattern Language (MPL) for encoding common game economy structures and design intent, and a Mechanics Design Assistant (MeDeA) for analyzing, explaining, understanding existing mechanics, and generating, filtering, exploring and applying design alternatives for modifying mechanics. As a result, game designers' productivity and game quality is increased by providing feedback and design alternatives early in the development cycle. Furthermore, the game economy modifications are applied at runtime using the MicroMachinations library, so that the effect of changes can be immediately experienced.

<p style="text-align: center; color: red;">**BONSAI Project-Team**</p>

7. New Results

7.1. Ancestral gene order reconstruction

In the field of **genomic rearrangement**, a topic of interest is to infer ancestral gene order from gene order known in extant species. The problem resumes to compute a set ancestral CARs (continuous ancestral regions) at a given node of a phylogeny. This work, initially published in a conference, was published this year in a journal [5].

7.2. Nonribosomal peptides

Norine is the unique and leading platform dedicated to computational biology analysis of nonribosomal peptides (NRPs). It is used by thousands of scientists all over the world to explore and better understand the diversity of the NRPs. To improve the data quality and quantity in Norine, we are now opening our resource to external contributors. To achieve this new challenge, we developed new tools (MyNorine, s2m) and communicate on our novelties.

- **Crowdsourcing.** To facilitate the submission of new nonribosomal peptides (NRPs) or modification of stored ones in Norine, we have developed a dedicated and user-friendly module named MyNorine [2]. It provides interactive forms to fill in the annotations with, for example, auto-completion and tools such as a monomeric structure editor. It has especially been designed for biologists and biochemists working on secondary metabolites to easily enrich the database with their own data.

- **Norine communication.** We advertise Norine by different promoting media. We organized an international workshop in Lille in October to teach biologists and biochemists how to annotate NRPs and their synthetases with bioinformatics tools such as Norine. It attracted 32 attendees from 8 countries. We participated, as invited contributors, to the special issue "Bioinformatics tools and approaches for synthetic biology" of the new journal Synthetic and Systems Biotechnology edited by KeAi Publishing, funded by Elsevier and Chinese Science Publishing & Media. Our article [6] describes the usefulness of Norine to discover novel nonribosomal peptides, with examples of biological results obtained thanks to Norine tools. More than 20 NRPs have already been submitted since September, proving the efficiency of our communication and usefulness and relevancy of Norine.

- **Monomeric structure.** The tool Smiles2Monomers (abbreviated s2m) infers efficiently and accurately the monomeric structure of a polymer from its chemical structure [1]. It is provided to the scientific community through the Norine website for on-line run or for download. Beside its utility to facilitate the annotation of new peptides, it allowed us to detect annotation errors in the Norine database.

7.3. High-throughput V(D)J repertoire analysis

High-throughput V(D)J repertoire analysis is an activity started in the group in 2012. As mentioned in previous reports, we produced a platform dedicated to analysing lymphocyte populations: Vidjil. Starting from DNA sequences, Vidjil is able to identify and quantify lymphocyte populations, visualise them and store metadata. Vidjil is now used routinely in Lille hospital and is also tested in other laboratories around the world.

With collaborators in Prague we used Vidjil in a retrospective study on patients suffering acute lymphoblastic leukemia [3]. The study identified a new measure of predicting relapse in patients, just a month after the diagnosis. This measure is simple as it relies on the diversity of the lymphocyte population.

7.4. Spaced seed coverage

In the field of spaced seed statistics these last two years, a new challenge is the selection of a set of spaced seeds that are at the same time sensitive, while providing a stable similarity measure for *alignment-free genomic sequence comparison*. One of the most stable estimators is the *coverage* provided by these seeds. We have proposed an efficient method to build the coverage automaton, in order to compute several statistics efficiently. This work was implemented in Iedera and published in the AISM journal [4].

7.5. Genome scaffolding with contaminated data

Scaffolding is a cornerstone in the assembly of genomes from next-generation sequencing data. It consists in ordering assembled sequences according to their putative order and orientation in the source genome. However, we are almost always in a setting where the genome is not known. Instead, order and orientation of sequences are inferred from partial information present in the sequencing data.

Unfortunately, sequencing data is noisy and often has contamination, i.e. a subset of the data which indicates a wrong genome order and/or orientation. We have investigated this effect and designed the first algorithm that explicitly models this contamination to better perform scaffolding.

This work appeared in the proceedings of the WABI 2015 conference [9] and has been accepted to the Bioinformatics journal, currently under revision. This work is in collaboration with K. Sahlin and L. Arvestad (KTH, Sweden).

7.6. Mining metatranscriptomic data

The team has recently developed the SortMeRNA software, which is a sequence analysis tool for filtering, mapping and OTU-picking NGS reads. The core algorithm is based on approximate seeds and allows for fast and sensitive analyses of nucleotide sequences. In [11], we demonstrate a computational technique for filtering ribosomal RNA from total RNA in metatranscriptomic data using it. Additionally, we propose a post-processing pipeline using the latest software tools to conduct further studies on the filtered data, including the reconstruction of mRNA transcripts for functional analyses and phylogenetic classification of a community using the ribosomal RNA. This work is a collaboration with Genoscope.

7.7. Structured RNAs

In many families of strutured RNAs, the signature of the family cannot be characterized by a single consensus structure, and is mainly described by a set of alternate secondary structures. For example, certain classes of RNAs adopt at least two distinct stable folding states to carry out their function. This is the case of riboswitches, that undergo structural changes upon binding with other molecules, and recently some other RNA regulators were proven to show evolutionary evidence for alternative structure. The necessity to take into account multiple structures also arises when modeling an RNA family with some structural variation across species, or when it comes to work with a set of predicted suboptimal foldings. In this perspective, we have introduced the concept of RNA multistructures, that is a formal grammar based framework specifically designed to model a set of alternate RNA secondary structures. Continuing our work of 2014, we provide several motivating examples and propose an efficient algorithm to search for RNA multistructures within a genomic sequence. This work was published in [7].

<p style="text-align:center">**DEFROST Team**</p>

7. New Results

7.1. Aggregate Constraints for Virtual Manipulation with Soft Fingers

In this work, we propose a new formulation of contact and friction laws, in the context of virtual graping. The work allows to reduce the number of contact and friction constraints, using volume interpenetration measure, instead of interpenetration distance. The work has been conducted in collaboration with Antony Talvas and Maud Machal (Inria Hybrid Team, Rennes) and Miguel Otaduy (URJC Madrid). It has been presented at the conference IEEE VR and published in the journal TVCG [5].

7.2. Haptic Rendering of Hyperelastic Models with Friction

We have reached an important milestone with this work: we have merge two important research tracks of these last years: On one hand, haptic rendering of friction contact between deformable objects ; on the ohter hand, real-time simulation of hyperelastic objects (particularly to simulate soft-tissues). This work has been conducted in collaboration with Hadrien Courtecuisse (Inria team Mimesis) and Hervé Delingette (Inria team Asclepios) [6]

7.3. Augmentation of Elastic Surfaces with Self-Occlusion Handling

In this work, we propose to recover the 3D shape and to augment elastic objects with self-occlusions handling, using only single view images. Shape recovery from a monocular video sequence is an underconstrained problem and many approaches have been proposed to enforce constraints and resolve the ambiguities. State-of-the art solutions enforce smoothness or geometric constraints, consider specific deformation properties such as inextensibility or resort to shading constraints. We propose a real-time method that uses a mechanical model and that is able to handle highly elastic objects. The problem is formulated as an energy minimization problem accounting for a non-linear elastic model constrained by external image points acquired from a monocular camera. This method prevents us from formulating restrictive assumptions and specific constraint terms in the minimization. In addition, we propose to handle self-occluded regions thanks to the ability of mechanical models to provide appropriate predictions of the shape. This result has been published in the journal TVCG [2] and has been extended to handle cutable objects and has been published as a SIGGRAPH poster [12].

7.4. Real-time control of soft-robots using asynchronous finite element modeling

Finite Element analysis can provide accurate deformable models for soft-robots. However, using such models is very difficult in a real-time system of control. In this study, we introduce a generic solution that enables a high-rate control and that is compatible with strong real-time constraints. From a Finite Element analysis, computed at low rate, an inverse model of the robot outputs the setpoint values for the actuator in order to obtain a desired trajectory. This inverse problem uses a QP (quadratic-programming) algorithm based on the equations set by the Finite Element Method. To improve the update rate performances, we propose an asynchronous simulation framework that provides a better trade-off between the deformation accuracy and the computational burden. Complex computations such as accurate FEM deformations are done at low frequency while the control is performed at high frequency with strong real-time constraints. The two simulation loops (high frequency and low frequency loops) are mechanically coupled in order to guarantee mechanical accuracy of the system over time. Finally, the validity of the multi-rate simulation is discussed based on measurements of the evolution in the QP matrix and an experimental validation is conducted to validate the correctness of the high-rate inverse model on a real robot. [8]

7.5. Domain decomposition approach for FEM quasistatic modeling and control of Continuum Robots with rigid vertebras

This study focuses on a new method dedicated to the modeling and control of Continuum Robots, based on the Finite Element Method (FEM) using quasi-static assumption. The modeling relies on a discretization of the continuum robots using 6 DoFs Frames along the structure of the robot that is compatible with the modeling of a sequence of rigid vertebras. When the robot's structure relies on rods with constant sections, internal forces are computed with beam elements, placed between two adjacent frames, that applies forces and torques. In the opposite, when the robot is composed of a complex shape deformable backbone separated by the rigid vertebras, a domain decomposition strategy is used to obtain an equivalent stiffness between two vertebras using volumetric FEM. In both cases, for solving the whole robot model and inverting it in real-time, the numerical method takes advantage of the serial nature of continuum robots, using a Block-Tri-Diagonal solver. The factor of improvement in the computation time reaches several order of magnitude compared to a classical FEM model, while keeping a good precision. The method has also been implemented and tested on a real pneumatic CBHA trunk designed by Festo Robotics and some complementarity examples have been generated numerically.[10]

<p style="text-align:center">**DOLPHIN Project-Team**</p>

7. New Results

7.1. Benchmarking Numerical Optimizers

Participants: D. Brockhoff, B. Derbel, A. Liefooghe, T.-D. Tran, D. Tušar, T. Tušar (DOLPHIN), O. Ait Elhara, A. Atamna, A. Auger, N. Hansen (TAO team, Inria Saclay), P. Preux (Univ. Lille 3), O. Mersmann, T. Wagner (TU Dortmund University, Germany), B. Bischl (LMU Munich, Germany), Y. Akimoto (Shinshu University, Japan)

In terms of benchmarking numerical optimization algorithms, our research effort went into two different directions. On the one hand, we continued our work on benchmarking single-objective optimization algorithms via the Coco platform in which we started to focus on algorithms for expensive optimization (problems for which only a few function evaluations are affordable). In particular, we benchmarked algorithm variants from the MATSuMoTo library [52], [50] and from the bandits-based global optimizer SOO (Simultaneous optimistic optimization) [33], and organized two workshops at CEC 2015 and GECCO 2015 (see also http://coco.gforge.inria.fr/). On the other hand, we started to develop an extension of the Coco platform towards multiobjective optimization and tried to establish the state of the art in single-objective benchmarking (target-based runtimes, data profiles, ...) also in the multi-objective case [30]. At the same time, we proposed a new bi-objective test suite, consisting of 300 well-understood, scalable test problems.

7.2. Handling numeric and temporal data in a local search-based classification algorithm

Participants: J. Jacques, L. Jourdan, C. Dhaenens, M. Vandomme

MOCA-I [20] is a highly efficient classification algorithm, primarily designed for knowledge extraction on large-scale, real-life medical data. This algorithm has been first extended to deal with numeric data [58], [46], through the definition of a model for classification rules on numeric attributes. Several neighborhood operators have been proposed, and compared, as components of the overarching local search metaheuristic guiding the discovery and optimization of these rules. A new model has also been proposed to handle temporal data. This model allows for the inclusion of sequences of events in classification rules, in addition to non-temporal attributes, thus building more informative classifiers. This model, along with various optimizations in the local search process, has been favorably compared to the previous MOCA-I algorithm and other standard classification algorithms. It is now used on real hospital data in order to evaluate its performance in a real environment.

7.3. MO-DYNAMOP

Participants: S. Jacquin, L. Jourdan, E-G. Talbi

The proposed method, MO-DYNAMOP generalize to multi-objective optimization, DYNAMOP, a state of the art optimizer, which was successfully applied to several MO problems. The specificity of this method is to combine aMO dynamic programming (MO-DP) with a MO evolutionary algorithm (MOEA). MO-DYNAMOP is applied to the first stage of the MO-UCP problem including minimization of gas emission. Since the second stage of the problem is now multi-objective, each solution of the first stage problem induces an entire Pareto front of the second stage problem. MO-UCP is solved by assigning an approximation of this Pareto front to each solution of the first stage problem. A comparison study with methods previously proposed in literature is performed. Experiments indicate that MO-DYNAMOP performs considerably better.

7.4. Decomposition-based multi-objective optimization

Participants: Dimo Brockhoff, Bilel Derbel, Arnaud Liefooghe, Gauvain Marquet, El-Ghazali Talbi, Saul Zapotecas-Martinez (external collaborators: Hernan Aguirre and Kiyoshi Tanaka, Shinshu Univ., Japan; Juan Palacios Alonso, Univ. Oviedo, Spain)

MOEA/D is an aggregation-based evolutionary algorithm which has been proved extremely efficient and effective for solving multi-objective optimization problems. It is based on the idea of decomposing the original multi-objective problem into several single-objective subproblems by means of well-defined scalarizing functions. Those single-objective subproblems are solved in a cooperative manner by defining a neighborhood relation between them. This makes MOEA/D particularly interesting when attempting to plug and to leverage single-objective optimizers in a multi-objective setting. For continuous optimization, we investigate in [49] the benefits that MOEA/D can achieve when coupled with CMA-ES, which is believed to be a powerful single-objective optimizer. We rely on the ability of CMA-ES to deal with injected solutions in order to update different covariance matrices with respect to each subproblem defined in MOEA/D. We show that by cooperatively evolving neighboring CMA-ES components, we are able to obtain competitive results for different multi-objective benchmark functions. Moreover, in the combinatorial case, we study in [48] the incorporation of geometric differential evolution (gDE), the discrete generalization of DE, into the MOEA/D framework. We conduct preliminary experiments in order to study the effectiveness of gDE when coupled with MOEA/D. Our results indicate that the proposed approach is highly competitive with respect to the original version of MOEA/D, when solving a combinatorial optimization problem having between two and four objective functions. In [36], we consider a bi-objective scheduling combinatorial problem in which task durations and due-dates are uncertain as a case study for MOEA/D. In particular, we investigate existing variants of MOEA/D and we propose a novel and simple alternative replacement component at the aim of maintaining population diversity. Through extensive experiments, we then provide a comprehensive analysis on the relative performance and the behavior of the considered algorithms. Besides being able to outperform existing MOEA/D variants, as well as the standard NSGA-II algorithm, our investigations provide new insights into the search ability of MOEA/D and highlight new research opportunities for improving its design components. At last, in [32], we propose the first large-scale message passing distributed scheme for parallelizing the computational flow of MOEA/D. We show how synchronicity and workload granularity can impact both quality and computing time, in an extremely fine-grained configuration. We deploy our distributed protocol using a large-scale environment of 128 computing cores. Besides being able to show significant speed-ups while maintaining competitive search quality, our experimental results provide insights into the behavior of the proposed scheme in terms of quality/speed-up trade-offs; thus pushing a step towards the achievement of effective and efficient parallel decomposition-based approaches for large-scale multi-objective optimization.

7.5. Fitness landscape analysis for multi-objective optimization

Participants: F. Daolio, A. Liefooghe (external collaborators: Sébastien Verel, Univ. Littoral Côte d'Opale, France; Hernan Aguirre and Kiyoshi Tanaka, Shinshu Univ., Japan)

Computationally hard multi-objective combinatorial optimization problems are common in practice, and numerous evolutionary multi-objective optimization (EMO) algorithms have been proposed to tackle them. Our aim is to understand which (and how) problem features impact the search performance of such approaches. In [38], we adopt a statistical approach, based on simple and multiple linear regression analysis, to enquire the expected running time of global SEMO with restart for identifying a $(1+\varepsilon)$-approximation of the Pareto set for small-size enumerable instances. Our analysis provides further insights on the EMO search behavior and on the most important features that characterize the difficulty of an instance for this class of problems and algorithms. In [31], we consider two prototypical dominance-based algorithms: a global EMO strategy using an ergodic variation operator (GSEMO) and a neighborhood-based local search heuristic (PLS). Their respective runtime is estimated on a benchmark of combinatorial problems with tunable ruggedness, objective space dimension, and objective correlation (ρMNK-landscapes). In other words, benchmark parameters define classes of instances with increasing empirical problem hardness; we enumerate and characterize the search space of small instances. Our study departs from simple performance comparison to systematically analyze

the correlations between runtime and problem features, contrasting their association with search performance within and across instance classes, for both chosen algorithms. A mixed-model approach then allows us to further generalize from the experimental design, supporting a sound assessment of the joint impact of instance features on EMO search performance. Next, in [28], we analyse the behavior and compares the performance of MOEA/D, IBEA using the binary additive epsilon and the hypervolume difference indicators, and AεSεH as representative algorithms of decomposition, indicators, and ε-dominance based approaches for many-objective optimization. We use small MNK-landscapes to trace the dynamics of the algorithms generating high-resolution approximations of the Pareto optimal set. Also, we use large MNK-landscapes to analyze their scalability to larger search spaces. At last, in [39], we report an experimental analysis on stochastic local search for approximating the Pareto set of bi-objective unconstrained binary quadratic programming problems. First, we investigate two scalarizing strategies that iteratively identify a high-quality solution for a sequence of sub-problems. Each sub-problem is based on a static or adaptive definition of weighted-sum aggregation coefficients, and is addressed by means of a state-of-the-art single-objective tabu search procedure. Next, we design a Pareto local search that iteratively improves a set of solutions based on a neighborhood structure and on the Pareto dominance relation. At last, we hybridize both classes of algorithms by combining a scalarizing and a Pareto local search in a sequential way. A comprehensive experimental analysis reveals the high performance of the proposed approaches, which substantially improve upon previous best-known solutions. Moreover, the obtained results show the superiority of the hybrid algorithm over non-hybrid ones in terms of solution quality, while requiring a competitive computational cost. In addition, a number of structural properties of the problem instances allow us to explain the main difficulties that the different classes of local search algorithms have to face.

7.6. Fitness Landscape of the Factoradic Representation on the PFSP

Participants: Marie-Eléonore Marmion (external collaborators: Olivier Regnier-Courdert, University of Aberdeen, UK)

Because permutation problems are particularly challenging to model and optimise, the possibility to represent solutions by means of factoradics has recently been investigated, allowing algorithms from other domains to be used. Initial results have shown that methods using factoradics can efficiently explore the search space, but also present difficulties to exploit the best areas. In [57], the fitness landscape of the factoradic representation and one of its simplest operator is studied on the Permutation Flowshop Scheduling Problem (PFSP). The analysis highlights the presence of many local optima and a high ruggedness, which confirms that the factoradic representations is not suited for local search. In addition, comparison with the classic permutation representation establishes that local moves on the factoradic representation are less able to lead to the global optima on the PFSP.

7.7. How Neutrality Helps Multiobjective Local Search Algorithms

Participants: Aymeric Blot, Clarisse Dhaenens, Laetitia Jourdan, Marie-Eléonore Marmion(external collaborators: Hernan Aguirre and Kiyoshi Tanaka, Shinshu Univ., Japan)

We extend the concept of neutrality used in single-objective optimization to the multi-objective context and investigate its effects on the performance of multi-objective dominance-based local search methods [29]. We discuss neutrality in single-objective optimization and fitness assignment in multi-objective algorithms to provide a general definition for neutrality applicable to multi-objective landscapes. We also put forward a definition of neutrality when Pareto dominance is used to compute fitness of solutions. Then, we focus on dedicated local search approaches that have shown good results in multi-objective combinatorial optimization. In such methods, particular attention is paid to the set of solutions selected for exploration, the way the neighborhood is explored, and how the candidate set to update the archive is defined. We investigate the last two of these three important steps from the perspective of neutrality in multi-objective landscapes, propose new strategies that take into account neutrality, and show that exploiting neutrality allows to improve the performance of dominance-based local search methods on bi-objective permutation flowshop scheduling problems. This work is a first step to integrate learning in strategies of local search algorithms.

7.8. Surrogate-assisted multiobjective evolutionary algorithm for fuzzy job shop problems

Participants: E-G. Talbi and Juan José Palacios, Jorge Puente, Camino R. Vela, Inés Gonzalez-Rodrıguez (Univ. Oviedo, Spain)

We have considered a job shop scheduling problem with uncertain processing times modelled as triangular fuzzy numbers and propose a multiobjective surrogate assisted evolutionary algorithm to optimise not only the schedule's fuzzy makespan but also the robustness of schedules with respect to different perturbations in the durations. The surrogate model is defined to avoid evaluating the robustness measure for some individuals and estimate it instead based on the robustness values of neighbouring individuals, where neighbour proximity is evaluated based on the similarity of fuzzy makespan values. The experimental results show that by using fitness estimation, it is possible to reach good fitness levels much faster than if all individuals are evaluated.

7.9. Bipartite matching approximation

Participants: F. Dufossé

Bipartite matching is a classical academic problem on bipartite graphs. Many iterative heuristics need an initial approximate matching with linear computational time. We have designed two randomized highly parallelizable algorithms with linear execution time and quality guarantee. The approximation guarantees have been proved to reach respectively 63demonstrate the speed-ups and validate the applicability and efficiency of these algorithms on general bipartite graphs. Comparisons with the more efficient suboptimal linear algorithms of bipartite matching demonstrate a lower efficiently in average but a similar execution time, and validate the quality guarantee on all experiments. This work has been published in [16].

7.10. Parallel B&B revisited for coprocessors using our new IVM data structure dedicated to permutation problems

Participants: J. Gmys, R. Leroy and N. Melab

This contribution is a joint work with M. Mezmaz and D. Tuyttens from University of Mons (UMONS).

Solving large permutation Combinatorial Optimization Problems (COPs) using Branch-and-Bound (B&B) algorithms results in the generation of a very large pool of subproblems. Therefore, defining a dedicated data structure is crucial to store and manage efficiently that pool. In the Ph.D thesis of R. Leroy [11], we have proposed an original data structure called Integer-Vector-Matrix (IVM) for permutation COPs based on the factorial number system. Consequently, we have redefined the operators of the B&B algorithm acting on it. For performance evaluation in terms of memory footprint and CPU time usage, we conduct a complexity analysis and an extensive experimentation using the permutation Flow-Shop Scheduling Problem (FSP) as a case study. Compared to the Head-Tail Linked List (LL) data structure often used for parallel B&B as in our work [11], IVM requires up to n times less memory than LL, n being the size of permutations. Moreover, the IVM-based B&B is up to one order of magnitude faster than its LL-based counterpart in managing the pool of subproblems. Another major contribution of this thesis is to revisit parallel B&B for multi-core processors and many-core coprocessors (GPU and MIC) using IVM and LL-based work stealing. Several challenging issues are addressed including work distribution using factoradic-based intervals on multi-core processors, thread/branch divergence and data placement optimization on GPU, and vectorization on Intel Xeon Phi. The contribution and some of its extensions have been published in [40], [18]. An extensive experimental study shows that the IVM-based approach outperforms its LL-based counterpart by a significant margin on multi-core processors as well as on coprocessors.

A major extension of this work has been proposed in [54] and awarded as a best paper consists in offloading all the operators of the B&B algorithm to the GPU. Four interval-based WS strategies have been investigated using IVM. An extensive experimentation allowed us to demontrate that the GPU-accelerated approach is 5 times faster than its multi-core counterpart.

7.11. large scale heterogeneous parallel B&B based on hybrid work-stealing

Participants: Bilel Derbel, Tuan Trong Vu

In [27], we investigate the design of parallel B&B in large scale heterogeneous compute environments where processing units can be composed of a mixture of multiple shared memory cores, multiple distributed CPUs and multiple GPUs devices. We describe two approaches based on hybrid work-stealing in shared and distributed memory systems, addressing the critical issue of how to map B&B workload with the different levels of parallelism exposed by the target compute platform. We also contribute a throughout large scale experimental study which allows us to derive a comprehensive and fair analysis of the proposed approaches under different system configurations using up to 16 GPUs and up to 512 distributed cores. Our results shed more light on the main challenges one has to face when tackling B&B algorithms while describing efficient techniques to address them. In particular, we are able to obtain linear speed-ups at moderate scales where adaptive load balancing among the heterogeneous compute resources is shown to have a significant impact on performance. At the largest scales, intra-node parallelism and hybrid decentralized load balancing is shown to have a crucial importance in order to alleviate locking issues among shared memory threads and to scale the distributed resources while optimizing communication costs and minimizing idle times.

7.12. A Multi-objective Evolutionary Algorithm for Cloud Platform Reconfiguration

Participants: F. Legillon, N. Melab and E-G. Talbi

This contribution published in [37] is a result of an industrial collaboration with Tasker Cloud services company.

Offers of public IAAS providers are dynamic: new providers enter the market, existing ones change their pricing or improve their offering. The decision on whether and how to improve already deployed platforms, either by reconfiguration or migration to another provider, can be modelled as an NP-hard optimization problem. In this paper, we define a new realistic model for this migration problem, based on a multi-objective Optimization formulation. An evolutionary approach is introduced to tackle the problem, using newly defined specific operators. Experiments are conducted on multiple realistic data-sets, showing that the evolutionary approach is viable to tackle real-size instances in a reasonable amount of time.

7.13. A multi-objective approach for energy-efficient scheduling of large workloads in multicore distributed systems

Participants: E-G. Talbi and B. Dorronsoro (Univ. Cadiz, Spain), S. Nesmachnow (Universidad de la República, Uruguay), J. Taheri, A. Zomaya (Univ. Sydney, Australia), P. Bouvry (Univ. Luxembourg

This work proposes a two-level strategy for scheduling large workloads of parallel applications in multicore distributed systems, taking into account the minimization of both the total computation time and the energy consumption of solutions. Nowadays, energy efficiency is of major concern when using large computing systems such as cluster, grid, and cloud computing facilities. In the approach proposed, a combination of higher-level (i.e., between distributed systems) and lower-level (i.e., within each data-center) schedulers are studied for finding efficient mappings of workflows into the resources in order to maximize the quality of service, while reducing the energy required to compute them. The experimental evaluation demonstrates that accurate schedules are computed by using combined list scheduling heuristics (accounting for both problem objectives) in the higher level, and ad-hoc scheduling techniques to take advantage of multicore infrastructures in the lower level. Solutions are also evaluated with two user- and administrator-oriented metrics. Significant improvements are reported on the two problem objectives when compared with traditional load-balancing and round-robin techniques [15].

DREAMPAL Project-Team

6. New Results

6.1. HoMade in 2015

6.1.1. Interruption support

In the last release of HoMade we introduced interruptions. Up to 7 interruptions are supported. The priority is static and each trap is associated to one of the 7 first VCs of the master, they are called trap1 .. trap7. Trap is par nature reflective. When a trap is raised the HoMade master reaches a no-preemptive kernel. Traps have no effect on the slaves, they can continue to work. At the end of trap execution, HoMade master resumes the sequential execution, trap codes should be clean and should restitute the stack as it was when they began. A WAIT instruction and a long IP cannot be interrupted. An example of interrupts is provided in the reconfiguration part later.

6.1.2. New assembly language

HoMade waits for two binary codes: one for the master and one for the slaves. These two codes are loaded via the UART port and triggers a global reset of all the softcores after. Binary codes are a sequence of 16 bits words finishing by a long word filled with 4 NULL. Our post fixed macro assembler generates some binary codes from text files. This assembly language introduces some flow controls like if for repeat. It is also based on PC and VC definitions. Now the particular operator := generates reflective behaviors via WIM instructions. The syntax is so simple than everybody can understand a program. A full new syntax description is available with the assembler on the official HoMade web site : https://sites.google.com/site/homadeguide/assembleur-homade-v6. Here is the code for a mono HoMade to implement a reflective execution of Fibonacci suite. Switches values are put on the top on the stack to indicate the position in the list we want to process. Different input buttons affect the execution: • Button 0 changes to soft fibo execution using some library IPs. SWAP ROT DUP = - + are IPs to change the tops of the stack or to process dyadic integer operators. • Button 1 changes to hard execution using fibo vhdl long IP • Other buttons process the current fibo (hard or soft).

```
  :IP fibo $AC54 ; // fibo hard IPcode 54
// XX = 1 YY = 1
program
  : read
    $1f // immediate hexa
    btnpush // IP reads buttons pushed
    switch // IP reads switches
  ;
  : fibo_soft // function declare
    1 1 rot
    3 -
    for
      dup rot +
    next
    swap
    drop
  ;
  VC fibo_dyn := fibo_soft // VC init soft
start
  begin
  read
```

```
      swap dup
      0 = // test button
      if // reflective process
        fibo_dyn := fibo_soft
      endif
      1 =
      if
        fibo_dyn := fibo
      endif
      fibo_dyn // call VC
      7seg // IP to print result
      $1f
      btn // button to pause
      7seg
      again // infinite loop
    endprogram
```
When the VC fibo_dyn is called, you call hard or soft Fibonacci version depending of the sequence of pushed button. The soft code is 7 time slower than the hard code. The extra cost due to reflective facility is 2 cycles by VC call.

6.1.3. Dynamic IP reconfiguration

Xilinx chips are offering capabilities to program some pre-reserved chip areas with different bitsreams and this during the execution itself. It is not instantaneous and even worse the reconfiguration time depends of the length of the bitstream (the size of the area). Do not abuse of partial reconfigurations! But for some applications where context evolves at a "human speed", our design can benefit of this functionality to adapt the hardware to the current context. It is easy to introduce this notion in HoMade: just insert an IP! This IP has to manage the bitstream memory and the ICAP to load them in the predefined areas. We develop a such IP for the master, without broadcast of bitstream to the slaves for the moment. This IP reconfiguration only needs to know the bitstream address. Effectively for Xilinx, the data inside the bitstream are sufficient to achieve the reconfiguration. We introduced the new keyword 'in the assembler in order to express IP reconfigurations. The declaration of reconfigurable IPs may also include the bitstream address. Now we can program dynamic partial reconfiguration of IPs using our dedicated IP that we developed. Furthermore we can couple the dynamic reconfiguration with the reflective notion. Here is a simple example with dynamic image filters. The filter processes 1 block of 3x3 pixels. The 9 pixels are stored on the 3 top of the stack by aggregation of 3 pixels per word. External actuators can change from one IP to the other. We used interrupts and traps to apply this migration.

```
  program // bistream addresses between ( )
    :IP IP_median $EC11 ($0);
    :IP IP_Sobel $EC22 ($49E);
    VC filter
    : T1
      IP_median ^^
      filter := IP_median
    ;
    : T2
      IP_moyenne ^^
      filter := IP_moyenne
      trap1 := T1 // interrupt level 1
      trap2 := T2 // Interrupt level 2
      : get3pix // must be defined &
    ;
  start
```

```
begin
  $7D for
    get3Pix // 3x3 pixels on stack
    get3Pix
    get3Pix
    -rot swap
    $7D for
      filter // current IP
      get3Pix // next 3 pixels
      -rot
    next
  next
  again
endprogram
```

Concerning dynamic reconfiguration of IPs, we are testing a dedicated IP to manage directly the ICAP of Xilinx. The different bitstreams are stored in DDR3 and this IP finds the starting address from the stack. Of course this is a long IP. Some optimization to broadcast efficiently the same bitstream towards different slave reconfigurable areas are still a big challenge with Xilinx architecture.

6.1.4. IP fusion

To be free from EDA companies, we are deploying IP fusion strategies to manage the dynamic reconfiguration by ourselves. We obtain good results concerning the reconfiguration time, but for large and very different IPs, the fusion works like an aggregation of two IPs and the surface gain is insignificant.

6.1.5. Using hardware parallelism for reducing power consumption in video streaming applications

In the PhD thesis of Karim Ali we exploited using a flexible parallel hardware-based architecture in conjunction with frequency scaling as a technique for reducing power consumption in video streaming applications. In this work, we derived equations to ease the calculation for the level of parallelism and the maximum depth for the FIFOs used for clock domain crossing. Accordingly, a design space was formed including all the design alternatives for the application. The preferable design alternative is selected in aware of how much hardware it costs and what power reduction goal it can satisfy. We used Xilinx Zynq ZC706 evaluation board to implement two video streaming applications: Video downscaler (1:16) and AES encryption algorithm to verify our approach. The experimental results showed up to 19.6% power reduction for the video downscaler and up to 5.4% for the AES encryption. The architecture and experimental results were published in a paper entitled "Using hardware parallelism for reducing power consumption in video streaming applications" at the 10th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC) in Jun 2015, Bremen, Germany [12].

In collaboration with NAVYA, we started the first steps to implement a stereo vision algorithm over a parallel architecture using FPGA technologies. The algorithm is based on a local approach for calculating the disparity map using sum of absolute difference between the right and the left image. As a first step, we exploited the possible optimization levels that can be applied at the software level. After that by using high level synthesis tool (Vivado HLS from Xilinx) the code was written in C in a way that facilitates its conversion into HDL files. Optimization techniques were applied to reduce both the hardware resources and time required for processing one frame. This design was tested experimentally to show around 50% decrease in the time required for processing one frame if compared to the software one. Currently, we are in the step of exploring more techniques for hardware optimization and decreasing the processing time to meet the industrial requirements of our partner.

6.1.6. *A scalable flexible and dynamic reconfigurable architecture for high performance embedded computing*

In collaboration with Nolam Embedded Systems (NES) and in the framework of the CIFRE PhD of Venkata-subramanian Viswanathan, we proposed a scalable and customizable reconfigurable computing platform, with a parallel full-duplex switched communication network, and a software execution model to redefine the computation, communication and reconfiguration paradigms in high performance embedded systems. High Performance Embedded Computing (HPEC) applications are becoming highly sophisticated and resource consuming for three reasons. First, they should capture and process real-time data from several I/O sources in parallel. Second, they should adapt their functionalities according to the application or environment variations within given Size Weight and Power (SWaP) constraints. Third, since they process several parallel I/O sources, applications are often distributed on multiple computing nodes making them highly parallel. Due to the hardware parallelism and I/O bandwidth offered by Field Programmable Gate Arrays (FPGAs), application can be duplicated several times to process parallel I/Os, making Single Program Multiple Data (SPMD) the favorite execution model for designers implementing parallel architectures on FPGAs. Furthermore Dynamic Partial Reconfiguration (DPR) feature allows efficient reuse of limited hardware resources, making FPGA a highly attractive solution for such applications. The problem with current HPEC systems is that, they are usually built to meet the needs of a specific application, i.e., lacks flexibility to upgrade the system or reuse existing hardware resources. On the other hand, applications that run on such hardware architectures are constantly being upgraded. Thus there is a real need for flexible and scalable hardware architectures and parallel execution models in order to easily upgrade the system and reuse hardware resources within acceptable time bounds. Thus these applications face challenges such as obsolescence, hardware redesign cost, sequential and slow reconfiguration, and wastage of computing power.

Addressing the challenges described above, we propose an architecture that allows the customization of computing nodes (FPGAs), broadcast of data (I/O, bitstreams) and reconfiguration several or a subset of computing nodes in parallel. The software environment leverages the potential of the hardware switch, to provide support for the SPMD execution model. Finally, in order to demonstrate the benefits of our architecture, we have implemented a scalable distributed secure H.264 encoding application along with several avionic communication protocols for data and control transfers between the nodes. We have used a FMC based high-speed serial Front Panel Data Port (sFPDP) data acquisition protocol to capture, encode and encrypt RAW video streams. The system has been implemented on 3 different FPGAs, respecting the SPMD execution model. In addition, we have also implemented modular I/Os by swapping I/O protocols dynamically when required by the system. We have thus demonstrated a scalable and flexible architecture and a parallel runtime reconfiguration model in order to manage several parallel input video sources. These results represent a conceptual proof of a massively parallel dynamically reconfigurable next generation embedded computers [16] [15]. The PhD of Venkatasubramanian Viswanathan has been defended in the 12th of october 2015 .

6.2. Language-Parametric Formal Methods

The HoMade assembly language is still evolving. Thus, our research in formal methods for programming languages kept the language-parametric nature that we decided upon when we started the project. The techniques and tools developed here will be instantiated on the HoMade assembly when it stabilizes. Our results are also applicable to general programming languages in order to target a broader audience.

In 2015 we have consolidated the results obtained in previous years, by making them more generally available and publishing them in high-end venues.

6.2.1. *Language Definitions as Rewrite Theories*

In [9] we study the foundations of $\mathbb{K}$, a formal framework for defining operational semantics of programming languages. The $\mathbb{K}$-Maude compiler translates $\mathbb{K}$ language definitions to Maude rewrite theories. The compiler enables program execution by using the Maude rewrite engine with the compiled definitions, and program analysis by using various Maude analysis tools. $\mathbb{K}$ supports symbolic execution in Maude by means of an

automatic transformation of language definitions. The transformed definition is called the *symbolic extension* of the original definition. In this paper we investigate the theoretical relationship between $\mathbb{K}$ language definitions and their Maude translations, between symbolic extensions of $\mathbb{K}$ definitions and their Maude translations, and how the relationship between $\mathbb{K}$ definitions and their symbolic extensions is reflected on their respective representations in Maude. In particular, the results show how analysis performed with Maude tools can be formally lifted up to the original language definitions.

6.2.2. *A Generic Framework for Symbolic Execution: Theory and Applications*

In [10], [17] we propose a language-independent symbolic execution framework. The approach is parameterised by a language definition, which consists of a signature for the language's syntax and execution infrastructure, a model interpreting the signature, and rewrite rules for the language's operational semantics. Then, symbolic execution amounts to computing symbolic paths using a *derivative* operation. We prove that the symbolic execution thus defined has the properties naturally expected from it, meaning that the feasible symbolic executions of a program and the concrete executions of the same program mutually simulate each other. We also show how a coinduction-based extension of symbolic execution can be used for the deductive verification of programs. We show how the proposed symbolic-execution approach, and the coinductive verification technique based on it, can be seamlessly implemented in language definition frameworks based on rewriting such as the $\mathbb{K}$ framework. A prototype implementation of our approach has been developed in $\mathbb{K}$. We illustrate it on the symbolic analysis and deductive verification of nontrivial programs.

6.2.3. *Symbolic Execution by Language Transformation*

In [2] we propose a language-independent symbolic execution framework for languages endowed with a formal operational semantics based on term rewriting. Starting from a given definition of a language, a new language definition is generated, with the same syntax as the original one, but whose semantical rules are transformed in order to rewrite over logical formulas denoting possibly infinite sets of program states. Then, the symbolic execution of concrete programs is, by definition, the execution of the same programs with the symbolic semantics. We prove that the symbolic execution thus defined has the properties naturally expected from it (with respect to concrete program execution). A prototype implementation of our approach was developed in the K Framework. We demonstrate the tool's genericity by instantiating it on several languages, and illustrate it on the reachability analysis and model checking of several programs.

6.2.4. *Program Equivalence by Circular Reasoning*

In [7] we propose a logic and a deductive system for stating and automatically proving the equivalence of programs written in languages having a rewriting-based operational semantics. The chosen equivalence is parametric in a so-called observation relation, and it says that two programs satisfying the observation relation will inevitably be, in the future, in the observation relation again. This notion of equivalence generalises several well-known equivalences and is appropriate for deterministic (or, at least, for confluent) programs. The deductive system is circular in nature and is proved sound and weakly complete; together, these results say that, when it terminates, our system correctly solves the given program-equivalence problem. We show that our approach is suitable for proving equivalence for terminating and non-terminating programs as well as for concrete and symbolic programs. The latter are programs in which some statements or expressions are symbolic variables. By proving the equivalence between symbolic programs, one proves the equivalence of (infinitely) many concrete programs obtained by replacing the variables by concrete statements or expressions. The approach is illustrated by proving program equivalence in two languages from different programming paradigms. The examples in the paper, as well as other examples, can be checked using an online tool.

6.2.5. *Verifying Reachability-Logic Properties on Rewriting-Logic Specifications*

Rewriting Logic is a simply, flexible, and powerful framework for specifying and analysing concurrent systems. Reachability Logic is a recently introduced formalism, which is currently used for defining the operational semantics of programming languages and for stating properties about program executions. Reachability Logic has its roots in a wider-spectrum framework, namely, in Rewriting Logic Semantics. In the invited

paper [10] we show how Reachability Logic can be adapted for stating properties of transition systems described by Rewriting-Logic specifications. We propose a procedure for verifying Rewriting-Logic specifications against Reachability-Logic properties. We prove the soundness of the procedure and illustrate it by verifying a communication protocol specified in Maude.

6.2.6. A Theoretical Foundation for Programming Languages Aggregation

This work was published as [11]. Programming languages should be formally specified in order to reason about programs written in them. We show that, given two formally specified programming languages, it is possible to construct the formal semantics of an aggregated language, in which programs consist of pairs of programs from the initial languages. The construction is based on algebraic techniques and it can be used to reduce relational properties (such as equivalence of programs) to reachability properties (in the aggregated language).

6.3. The SCAC Model : a weakly-coupled execution model for MPSoC

Synchronous Communication Asynchronous Computation (SCAC) is an execution model that separates the execution of communication phases from those of computation in order to facilitate their overlapping, thus covering the data transfer time. To allow the simultaneous execution of these two phases, we propose an approach based on three levels : two globally-centralized/locally-distributed hierarchical control levels and a parallel computation level.

G-MPSoC [5] is a SCAC System-on-Chip implementation based on a grid of clusters of Hardware and Software Computation Elements with different size, performance, and complexity. It is composed of parametric IP-reused modules: processor, controller, accelerator, memory, interconnection network, etc. to build different architecture configurations. The generic structure of G-MPSoC facilitates its adaptation to the intensive signal processing applications requirements.

The communication phase in SCAC System-on-Chip should be as fast as possible to avoid compromising parallel computing, using small and low power consumption modules to facilitate the interconnection network extensibility in a scalable system. To meet these criteria and based on a module reuse methodology, we chose to integrate a reconfigurable SCAC-Net [14] interconnection network to communicate data in our system. The SCAC-Net network is composed of communication modules as the number of the nodes used by the system. Using generic parameters, the topology of SCAC-Net network can be easily configured according to the needed communication which give more flexibility to the system.

<div align="center" style="color:red">

FUN Project-Team

</div>

7. New Results

7.1. Data gathering and coverage in WSN

Participants: Nathalie Mitton, Tahiry Razafindralambo, Arunabha Sen.

Data availability is one of the main goals and challenges in Future Ubiquitous Network and especially in Wireless Sensor Networks. Indeed, gathering and collecting data in a mobile environment is a very challenging task. In [12], the authors uses data mules to organize the data collection from the sensor in the field. The results presented in [12] are based on some new and unique assumptions. First, it is assumed that the mules are mobile but also the sensors that generate the data to be collected. Second, we collection time is not the first optimization criteria. The focus of the paper is on minimizing the number of mules given a time contraints. The problem is shown to be NP complete and a transformation of the problem into a minimum flow problem allow the computation of optimal solution using Integer Linear Programming.

The results presented in [5] use the assumption of a mobile objects tracked by some other mobile objects as in [12]. In the case of [5] the focus in on coverage of mobile targets by mobile Unmanned Aerial Vehicle. The paper takes two major assumptions regarding the limited energy of UAV and the observation range. These two constraints are linked with each other since when the UAV increase its altitude, it consumes more energy but also increase its observation range. The problem under consideration is mathematically represented by defining mixed integer non-linear optimization models. Heuristic procedures are defined and they are based on restricted mixed integer programming (MIP) formulation of the problem. A computational study is carried out to assess the behavior of the proposed models and MIP-based heuristics.

7.2. Routing in FUN

Participants: Nathalie Mitton, Mouna Rekik.

Geographic routing is an attractive routing strategy in wireless sensor networks. It works well in dense networks, but it may suffer from the void problem. For this purpose, a recovery step is required to guarantee packet delivery. Face routing has widely been used as a recovery strategy since proved to guarantee delivery. However, it relies on a planar graph not always achievable in realistic wireless networks and may generate long paths. In [25], we propose GRACO, a new geographic routing algorithm that combines a greedy forwarding and a recovery strategy based on swarm intelligence. During recovery, ant packets search for alternative paths and drop pheromone trails to guide next packets within the network. GRACO avoids holes and produces near optimal paths. Simulation results demonstrate that GRACO leads to a significant improvement of routing performance and scalability when compared to the literature algorithms.

GRACO has first been designed in the general case. We then studied its applicability to the Virtual Power Plants and their specific data packets with different priorities [22], [24]. Indeed, the Smart Grid (SG) incorporates communication networks to the conventional electricity system in order to intelligently integrate distributed energy resources (DERs) and allow for demand side management. The move to Smart grid in developing countries has to cope with great disparities of ICT infrastructures even within the same city. Besides, individual DERs are often too small to be allowed access to energy market, likewise power utilities are unable to effectively control and manage small DERs. We propose the use of affordable and scalable wireless communication technology to aggregate geographically sparse DERs into a single virtual power plant. The enrollment of prosumers in the VPP is conditional to financial performance of the plant. Thus, the VPPs are dynamic and are expected to scale up as more and more prosumers are attracted by their financial benefits. the communication network has to follow this progression and therefore to be scalable and rapidly deploy-able. We present a routing algorithm for data communication within the VPP to support centralized, decentralized or fully distributed control of the VPP's DERs.

Based on this study, we adapted GRACO so it can fit the specific cases of Smart Grid [23] and more specifically to the Neighbor Area Networks (NAN) of Smart Grids, or distribution segment of the power system in the smart grid (SG). The deployment of ICT to support conventional grid will solve legacy problems that used to prevent implementation of smart services such as smart metering, demand side management or the integration of Distributed Energy Resources (DERs) within the smart grid. We demonstrate the effectiveness of GRACO in terms of scalability, peer-to-peer routing, end-to-end delay and delivery rate.

7.3. Deployment and Self-Deployment in FUN

Participants: Nathalie Mitton, Valeria Loscri, Tahiry Razafindralambo.

Mobility management is a difficult task in autonomous networks. However, mobility provide a huge advantage when in comes to specific scenario such as emergency-related ones especially when network connection must be restored to provide basic network access to users. [3] investigates the potential of spontaneous networks for providing Internet connectivity over the emergency area through the sharing of resources owned by the end-user devices. Novel and extremely flexible network deployment strategies are required in order to cope with the user mobility, the limited communication capabilities of wireless devices, and the intrinsic dynamics of traffic loads and QoS requirements. In [3], a novel architecture is proposed to take advantage of existing end-user devices and some algorithm, are described to build and efficiently exploit the spontaneous emergency network.

Following the emergency scenario described in [3], [15] and [21] describe an algorithm to minimizes the control trafic generated by specific nodes in the network used repair the network and the deployment of these specific nodes, This nodes, forming a substitution network, in case of emergency, are injected autonomously in the network by the network to restore basic network service. In order to increase the performance of the network, the injected nodes called substitution routers, use their ability to move to change the shape of the network and to increase its performance. These movements needs huge amount of control messages to maintain consistency regarding routers' positions. [15] and [21] give an algorithm for the deployment of these routers and the autoregressive time serie model to reduce the amount of control trafic used for the deployment.

7.4. Smart Cities

Participants: Nathalie Mitton, Valeria Loscri, Riccardo Petrolo.

Smart City represents one of the most promising, prominent and challenging Internet of Things (IoT) applications, but recent ICT trends suggest more and more that cities could also benefit from Cloud computing. The convergence of IoT paradigm and Cloud computing technology, can play a fundamental role for developing of highly level and organized cities form an ICT point of view, but it is of paramount importance to deal a critical analysis to identify the issues and challenges deriving from this synergy. This detailed study has been dealt in [7], where it is shown as the semantic annotation of the sensors in the cloud, and innovative services can be implemented and considered by bridging Cloud and Internet of Things. The Cloud of Things (CoT) paradigm is also considered in [16], where it is shown how the CoT arrives to better distribute resources, putting together and enabling therefore a horizontal integration of various Internet of Things(IoT) platforms. Semantic interoperability of diverse IoT platforms are also a key concept in [18], where the virtualization of different IoT systems in order to model and represent the architecture in accordance with the common standards-based IoT ontologies is applied. The environment comes with a range of visual drag-and-drop tools, which boosts developers' productivity.

7.5. RFID

Participant: Nathalie Mitton.

One of the devices under consideration by the FUN team is RFID. One of the main issues to widely deploy RFID reader is reader-to-reader collision. Indeed, when the electromagnetic fields of the readers overlap, a collision occurs on the tag laying in the overlapping section and cannot be read. In [10], we propose a high adaptive contention-based medium access control (HAMAC) protocol that considerably reduces readers collision problems in a large-scale dynamic RFID system. HAMAC is based only on realistic assumptions that can be experimented and does not require any additional components on RFID reader in order to improve the performance in terms of throughput, fairness and latency. The central idea of the HAMAC is for the RFID reader to use a WSN-like CSMA approach and to set its initial backoff counter to the maximum value that allows to mitigate collision. Then, according to the network congestion on physical channels the reader tries to dynamically control its contention window by linear decreasing on selected physical channel or multiplicative decreasing after scanning all available physical channels. Extensive simulations are proposed to highlight the performance of HAMAC compared to literature's work in large-scale RFID systems where both readers and tags are mobile. Simulation results show the effectiveness and robustness of the proposed anti-collision protocol in terms of network throughput, fairness, coverage and time to read all tags.

7.6. Localization

Participants: Nathalie Mitton, Roudy Dagher, Valeria Loscri, Salvatore Guzzo Bonifacio.

[20] presents our approach to localize a node with the use of only one landmark. It is a passive and non intrusive cross-layer approach that relies on a signal processing of all received signals Results are evaluated by simulation and show good accuracy. To complete the previous study, we developed [11] a novel array-based method to estimate the path loss exponent (PLE). The method is designed as a part of an automatic calibration step, prior to localization of a source transmitting in the near-far field of the array. The method only requires the knowledge of the ranges between the array elements. By making the antenna elements transmit in turn, the array response model in the near-far field is exploited to estimate the current environment PLE. Simulation results show that this method can achieve good performance with one transmission round. The performance of the PLE estimation is investigated in the context of source localization with a sensitivity analysis to the PLE estimation. These works are the purpose of a pending patent (submitted in March 2015).

Alternatively, we derive similar localization schemes to enable a cooperation between mobile robots to locale a target based on RSSI [13]. Received Signal Strength Indicator (RSSI) is commonly considered and is very popular for target localization applications, since it does not require extra-circuitry and is always available on current devices. Unfortunately, target localizations based on RSSI are affected with many issues, above all in indoor environments. In this paper, we focus on the pervasive localization of target objects in an unknown environment. In order to accomplish the localization task, we implement an Associative Search Network (ASN) on the robots and we deploy a real test-bed to evaluate the effectiveness of the ASN for target localization. The ASN is based on the computation of weights, to "dictate" the correct direction of movement, closer to the target. Results show that RSSI through an ASN is effective to localize a target, since there is an implicit mechanism of correction, deriving from the learning approach implemented in the ASN.

7.7. Vehicular Networks

Participants: Nathalie Mitton, Valeria Loscri.

In the framework of our collaboration with Southern University in China, we investigate a specific issue in Vehicular AdHoc Networks (VANET), the information delivery delay analysis for roadside unit deployment in a VANET with intermittent connectivity [9]. A mathematical model is developed to describe the relationship between the average delay for delivering road condition information and the distance between two neighbor RSUs deployed along a road. The derived mathematical model considers a straight highway scenario where two RSUs are deployed at a distance without any direct connection and vehicles are sparsely distributed on the road with road condition information randomly generated between the two neighbor RSUs. Moreover, the model takes into account the vehicle speed, the vehicle density, the likelihood of an incident, and the distance between two RSUs. The effectiveness of the derived mathematical model is verified through simulation results.

Given the information delivery delay constraint of a time-critical application, this model can be used to estimate the maximum distance allowed between two neighbor RSUs, which can provide a reference for the deployment of RSUs in such scenarios.

But Vehicular Networks can also convey social networks. In [30], we survey recent literature on Vehicular Social Networks that are a particular class of vehicular ad hoc networks, characterized by social aspects and features. Starting from this pillar, we investigate perspectives of next generation vehicles under the assumption of social networking for vehicular applications (i.e., safety and entertainment applications). This paper plays a role as a starting point about socially-inspired vehicles, and main related applications, as well as communication techniques. Vehicular communications can be considered as the "first social network for automobiles", since each driver can share data with other neighbors. As an instance, heavy traffic is a common occurrence in some areas on the roads (e.g., at intersections, taxi loading/unloading areas, and so on); as a consequence, roads become a popular social place for vehicles to connect to each other. Human factors are then involved in vehicular ad hoc networks, not only due to the safety related applications, but also for entertainment purpose. Social characteristics and human behavior largely impact on vehicular ad hoc networks, and this arises to the vehicular social networks, which are formed when vehicles (individuals) "socialize" and share common interests. This survey describes the main features of vehicular social networks, from novel emerging technologies to social aspects used for mobile applications, as well as main issues and challenges. Vehicular social networks are described as decentralized opportunistic communication networks formed among vehicles. They exploit mobility aspects, and basics of traditional social networks, in order to create novel approaches of message exchange through the detection of dynamic social structures. An overview of the main state-of-the-art on safety and entertainment applications relying on social networking solutions is also provided.

7.8. FIT

Participants: Nathalie Mitton, Julien Vandaele.

The universal proliferation of intelligent objects is making Internet of Things (IoT) a reality; to operate on a large scale it will critically rely on new, seamless, forms of communications. But how can innovations be validated in a controlled environment, before being massively deployed into the real world? Several platforms have been deployed to address this issue. In [8], we browse a survey of them, highlighting their characteristics and given some tips to choose the most appropriate to our needs.

Our team has contributed to the deployment of the FIT IoT-LAB platform [2], [19], [27], which addresses this challenge by offering a unique open first class service to all IoT developers, researchers, integrators and developers: a large-scale experimental testbed allowing design, development, deployment and testing of innovative IoT applications, in order to test the future and make it safe. One of the specific deployment focuses on the automatic docking of robots for energy recharge. We explain it in [17]. The objective is to achieve long-term autonomous robots within an experiment test-bed. We propose to combine the use of QR codes as landmarks and Infrared distance sensors. The relative size of the lateral edges of the visual pattern is used to position the robot in relation with the dock. Infrared distance sensors are then used to perform different approaching strategies depending on the distance. Experiments show that the proposed solution is fully operational and robust. Not to rely exclusively on visual pattern recognition avoids potential errors induced by camera calibration. Additionally, as a positive side effect, the use of Infrared sensors allows the robot to avoid obstacles while docking. The finality of such an approach is to integrate these robots into the FIT IoT Lab experimental testbed which allows any experimenter to book wireless resources such as wireless sensors remotely and to test their own code. Wifibots holding wireless sensors will be integrated as additional reservable resources of the platform to enlarge the set of possible experimentations with mobile entities.

7.9. New and other communication paradigms

Participants: Nathalie Mitton, Valeria Loscri, Arash Maskooki, Gabriele Sabatino.

Interconnection and self-organized systems are normally populated with heterogeneous and different devices. The differences range from computational capabilities, storage size, etc. Instead of considering the heterogeneity as a limitation, it is possible to "turn it"as a primitive control of the system, in order to realize more robust and more resilient communication systems. Based on these premises, we identify specific situations, where mobile nodes with a plethora of interesting features and sensing capacities, can be exploited by configuring them in such a way to make them playing different roles in respect of them for which they have been initially conceived [4]. The differentiated use of devices, together with a careful analysis of the characteristics and performance requirements of the current and the future networks, allow the adaptation to the exponential growth in demand for high bandwidth applications [26]. This is exactly the philosophy embraced in [28], where Software Defined Radio (SDR) and Cognitive Radio (CR) have been considered and analyzed in a novel context, namely body networked systems. A detailed analysis of body systems as networked systems has also been considered in [6] and [14]. In [6] a novel communication paradigm, namely a molecular communication, has been considered to show how a nanoparticulate system can be suitable to coexist in a biological environment. An experimental analysis to asses the theoretical assumption has been developed in [14]. In order to asses new/alternative communication paradigms, there is the necessity from one side to consider and analyze the specific context and its level of interaction with the communication system and on the other side the correct identification of the specific features of the communication paradigm itself. This type of analysis allowed the design and implementation of an acoustic communication approach [29], where the ultrasound represent the wave carriers of data information. This "unusual" transmission means has been selected as the most suitable in a context as the body, where the aqueous environment makes it not suitable for more "traditional" communication paradigms, e.g. the one based on Radio Frequency (RF) waves.

INOCS Team (section vide)

LINKS Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Recursive queries

P. Bourhis published a paper at IJCAI [17] in cooperation with the University of Dresden in Germany. There he developed a highly expressive Web query language of the Datalog family, for which static analysis problems such as query containment remain decidable.

In cooperation with Links' associated team in Oxford, P. Bourhis obtained an article at ACM TODS [5], where he studies the access of hidden data by recursive queries.

V. Hugot, A. Boiret, and J. Niehren study monadic second-order logic for unordered trees with data constraints on siblings. This language can be used to define recursive queries and schemas on unordered data trees [13]. They study restrictions of the logics, for which the usual static analysis problems become decidable, and study the complexity of the decidable cases. This work was done in cooperation with Paris 7.

7.1.2. Schemas

I. Boneva and S. Staworko contribute at ICDT the RDF schema language SheX [22], which they developed in cooperation with members of the W3C. The usual open world approach of RDF is schemaless in that the alphabets of RDF data are left open, so that data from different sources and with different alphabets can be unified. This raises serious problems for query writing and thus for linked data integration, since a query may become invalid when the alphabet changes. A SheX schema allows to express constraints on the alphabets, node labels and edge labels of RDF graphs, so that database queries become safe with respect to future changes without closing the alphabet. In a previous work the studied the case of XML data trees instead of RDF graphs [6].

A. Lemay and J. Niehren propose sublinear algorithms in the style of probabilistic property testing for validating XML data trees with respect to DTD [20].

P. Bourhis studies streaming bounded repair with respect to schema violations [8]. This work is done in a cooperation with the University of Bordeaux and the University of Santiago in Chile.

7.1.3. Provenance

P. Bourhis obtained an ICALP paper [11] in cooperation with Télécom ParisTech. They show how to propagate provenance information for monadic second-order logics on trees or tree like structures with polynomial data complexity. In their provenance framework, they can show how to generalize various aggregation tasks for monadic second-order logics, that were known to be solvable with polynomal data complexity before.

In a cooperation with Tel Aviv, P. Bourhis obtained a ACM CIKM paper [18], where they show how to approximately summarize data provenance.

7.1.4. Data integration

In a cooperation with the University of Toronto, R. Ciucanu obtained a paper at PVLDB [4] on how to gain control over data integration evaluations. I. Boneva, A. Bonifati and R. Ciucaniu presented a paper on graph data exchange with target constraints [14] in the GraphQ workshop, and proved that query answering is intractable in this context.

7.2. Managing Dynamic Linked Data

7.2.1. Complex event processing

T. Sebastian, J. Niehren and D. Debarbieux propose early nested word automata for evaluating navigational XPath queries on XML streams [9]. They show how to approximate earliest query answering for such queries in a highly efficient manner and with very good precision in practice, while exact earliest query answering is known to untractable for XPath. This work was done in an industrial cooparation with Innovimax from Paris and in cooperation with the University of Bordeaux. In a follow-up work [21] they show that the XPath streaming algorithm for early nested word automata can be speed up considerably, when combining it with projection algorithms for nested word automata that they developed.

J. Niehren developed X-Fun [19] a uniform programming language for implementing XML standards, and showed how to implement XSLT, XProc, and XSLT in this manner. This work, that is fully implemented, was done in cooperation with the University of Bratislava.

7.2.2. Data-centered workflows

P. Bourhis presents highly expressive query languages as needed for data-centric workflows in the context of Active XML [3] in cooperation with the Dahu project from Inria Saclay.

J. Niehren presents a general framework for the reasoning with observational program semantics [10] in a cooperation with the Universities of Frankfurt and Saarbrücken in Germany.

7.3. Linking Data Graphs

S. Staworko obtained his HDR for his work on symbolic inference methods for databases [2]. R. Ciucanu obtained his PhD for his work on cross-model query inference [1] supervised bu A. Bonifati.

7.3.1. Learning path queries

A. Lemay, R. Ciucanu, and A. Bonifat have a paper and a demo at EDBT showing how to learn simple path queries on graph databases based on automata techniques [16], [15], [25], [24]. This is a very interesting starting point for using automata inference techniques in the context of graph databases.

S. Staworko obtained a paper at ICDT where he shows how to infer XML Twig queries from examples [23]. This work is done in cooperation with the University of Wrazlaw.

7.3.2. Learning join queries

R. Ciucanu, A. Boneva, and S. Staworko published an ACM TODS article [7] showing how to learn join queries for relational databases from examples. This is the first query learning algorithm satisfying Gold's learning model, that relies on equalities of data values rather than on structural information.

MAGNET Team

7. New Results

7.1. Decentralized Estimation in Networks

In [3], we studied the problem of decentralized estimation in networks, where each node of the network holds a data point and the goal is to estimate some statistics on the entire data under communication constraints imposed by the graph topology of the network. This generic problem has many applications in Internet of Things as well as for extracting knowledge from massive information graphs such as interlinked Web documents and online social media. In this work, we focused on estimating pairwise mean statistics. Popular examples of such statistics include the sample variance, the average distance and the Area Under the ROC Curve, among others. We proposed new synchronous and asynchronous randomized gossip algorithms which simultaneously propagate data across the network and maintain local estimates of the quantity of interest. We establish convergence rate bounds of $O(1/t)$ and $O(\log t/t)$ for the synchronous and asynchronous cases respectively, where t is the number of iterations, with explicit data and network dependent terms. Beyond favorable comparisons in terms of rate analysis, numerical experiments provide empirical evidence the proposed algorithms surpasses the previously introduced approach.

7.2. Large-Scale Learning with Higher-Order Risk Functionals

In [6], we studied learning problems where the performance criterion consists of an average over tuples (e.g., pairs or triplets) of observations rather than over individual observations, as in many learning problems involving networked data (e.g., link prediction), but also in metric learning and ranking. In this setting, the empirical risk to be optimized takes the form of a U-statistic, and its terms are highly dependent and thus violate the classic i.i.d. assumption. In this work, we focused on how to best implement a stochastic approximation approach to solve such risk minimization problems in the large-scale setting. We argue that gradient estimates should be obtained by sampling tuples of data points with replacement (incomplete U-statistics) rather than sampling data points without replacement (complete U-statistics based on subsamples). We develop a theoretical framework accounting for the substantial impact of this strategy on the generalization ability of the prediction model returned by the Stochastic Gradient Descent (SGD) algorithm. It reveals that the method we promote achieves a much better trade-off between statistical accuracy and computational cost. Beyond the rate bound analysis, we provide strong empirical evidence of the superiority of the proposed approach on metric learning and ranking problems.

7.3. Natural Language Processing

In [4], we introduce a new structured model for learning anaphoricity detection and coreference resolution in a joint fashion. Specifically, we use a latent tree to represent the full coreference and anaphoric structure of a document at a global level, and we jointly learn the parameters of the two models using a version of the structured perceptron algorithm. Our joint structured model is further refined by the use of pairwise constraints which help the model to capture accurately certain patterns of coreference. Our experiments on the CoNLL-2012 English datasets show large improvements in both coreference resolution and anaphoricity detection, compared to various competing architectures. Our best coreference system obtains a CoNLL score of 81.97 on gold mentions, which is to date the best score reported on this setting.

In [2], we present a detailed comparative framework for assessing the usefulness of unsupervised word representations for identifying so-called implicit discourse relations. Specifically, we compare standard one-hot word pair representations against low-dimensional ones based on Brown clusters and word embeddings. We also consider various word vector combination schemes for deriving discourse segment representations from word vectors, and compare representations based either on all words or limited to head words. Our main finding is that denser representations systematically outperform sparser ones and give state-of-the-art performance or above without the need for additional hand-crafted features.

7.4. Some Ongoing Work

7.4.1. *Metric Learning for Graph-based Label Propagation*

The efficiency of graph-based semi-supervised algorithms depends on the graph of instances on which they are applied. The instances are often in a vectorial form before a graph linking them is built. The construction of the graph relies on a metric over the vectorial space that helps define the weight of the connection between entities. The typical choice for this metric is usually a distance or a similarity measure based on the Euclidean norm. We claim that in some cases the Euclidean norm on the initial vectorial space might not be the most appropriate to solve the task efficiently.

In a paper currently under review, we proposed an algorithm that aims at learning the most appropriate vectorial representation for building a graph on which label propagation is solved efficiently, with theoretical guarantees on the classification performance.

7.4.2. *Link Classification in Signed Graphs*

We worked on active link classification in signed graphs. Namely, the idea is to build a spanning tree of the graph and query all its edge signs. In the two clusters case, this allows to predict the sign of an edge between nodes u and v as the product of the signs of edge along the path in the spanning tree from u to v. It turns out that ensuring low error rate amounts to minimizing the stretch, a long open standing problem known as Low Stretch Spanning Tree [11]. While we are still working on the theoretical analysis, experimental results showed that our construction is generally competitive with a simple yet efficient baseline and outperforms it for specific graph geometry like grid graphs.

Moreover, based on experimental observations, we will also analyze a heuristic which exhibits good performance at a very low computational cost and is therefore well suited for large-scale graphs. In a nutshell, it predicts the sign of an edge from u to v based on the fraction of u negative outgoing edges and v negative incoming edges, exploiting a behavioral consistency bias from signed social network users.

Going further in link classification, we believe that the notion of sign can be extended, going from one binary label per edge to a more holistic approach where the similarity between two nodes is measured across different contexts. These contexts are represented by vectors whose dimension matches the dimension of unknown feature vectors associated with each node. The goal is to answer queries of the form: how similar are nodes u and v along a specific context? We first plan to validate the relevance of this modeling on real-world problems, then test baseline methods on synthetic and real data before looking for a more effective, online prediction method.

7.4.3. *Graph-based Learning for Dependency Parsing*

We are investigating the use of different graph-based learning techniques such as k-nearest neighbors classification and label propagation for the problem of dependency parsing. While most of current approaches rely on learning a single scoring model (through SVM, MIRA, neural networks) from a large set of hand annotated training data (usually thousands of sentences), we are interested in using the sentence space geometry (approximated via a similarity graph over some labeled and unlabeled sentences) to tune the model to better fit a given sentence. This amounts to learning a slightly different model for each unlabeled sentence.

In order to successfully parse sentences in this setting, we need to propagate parsing information from labeled sentences to unlabeled ones through the graph. In order to build a similarity graph well suited to dependency parsing, we worked on learning a similarity function between pairs of sentences, based on the idea that two sentences are similar if they have similar parse trees. We will then investigate how to propagate the trees (which may be of varying sizes) through the graph and consider several propagation schemes.

<div align="center">

MEPHYSTO Team

</div>

7. New Results

7.1. Quantitative stochastic homogenization

7.1.1. Discrete equations

7.1.1.1. Decay of the semi-group

A. Gloria, S. Neukamm (Univ. Dresden), and F. Otto (MPI for mathematics in the sciences, Leipzig) developed in [15] a general approach to quantify ergodicity in stochastic homogenization of scalar discrete elliptic equations. Using a parabolic approach, they obtained optimal estimates on the time-decay of the so-called environment seen from the particle. This allowed them to prove optimal bounds on the corrector gradient and the corrector itself in any dimension (thus improving on [5]). They also obtained the first error analysis of the popular periodization method to approximate the homogenized coefficients.

7.1.1.2. Quantitative CLT

In [16], A. Gloria and J. Nolen (Duke Univ.) proved a quantitative central limit theorem for the effective conductance on the discrete torus. In particular, they quantified the Wasserstein distance between a normal random variable and the CLT-like rescaling of the difference between the approximation of the effective conductance by periodization and the effective conductance. Their estimate is sharp and shows that the Wasserstein distance goes to zero (up to logarithmic factors) as if the energy density of the corrector was iid (which it is not). This completes and settles the analysis started in [15] on the approximation of homogenized coefficients by periodization by characterizing the limiting law in addition to the scaling.

7.1.2. Continuum equations

7.1.2.1. Scalar equations with random coefficients

In [17], A. Gloria and F. Otto extended their results [4] , [5] on discrete elliptic equations to the continuum setting. They treated in addition the case of non-symmetric coefficients, and obtained optimal estimates in all dimensions by the elliptic approach (whereas [4],[5] were suboptimal for $d = 2$).

In [14], A. Gloria and D. Marahrens (MPI for mathematics in the sciences, Leipzig) extended the annealed results [14] on the discrete Green function by D. Marahrens and F. Otto to the continuum setting. As a by-product of their result, they obtained new results in uncertainty quantification by estimating optimally the variance of the solution of an elliptic PDE whose coefficients are perturbed by some noise with short range of dependence.

7.1.2.2. Systems with random coefficients

In a revised version of [58], A. Gloria, S. Neukamm, and F. Otto developed a regularity theory for random elliptic operators inspired by the contributions of Avellaneda and Lin [43] in the periodic setting and of S. Armstrong with C. Smart [42]. This allowed them to consider coefficients with arbitrarily slow decaying correlations in the form of a family of correlated Gaussian fields, and obtain (in the new version of this paper) a family of estimates with optimal rates and exponential-type integrability.

In [35], A. Gloria and F. Otto obtained the first nearly-optimal estimates with optimal stochastic integrability on the corrector for linear elliptic systems whose coefficients satisfy a finite range of dependence assumption (thus avoiding the functional inequalities they considered so far).

7.1.2.3. Systems with almost periodic coefficients

In [23], S. Armstrong, A. Gloria and T. Kuusi (Aalto University) obtained the first improvement over the thirty year-old result by Kozlov [60] on almost periodic homogenization. In particular they introduced a class of almost periodic coefficients which are not quasi-periodic (and thus strictly contains the Kozlov class) and for which almost periodic correctors exist. Their approach combines the regularity theory developed by S. Armstrong and C. Smart in [42] and adapted to the almost periodic setting by S. Armstrong and Z. Shen [41], a new quantification of almost-periodicity, and a sensitivity calculus in the spirit of [4].

7.1.3. Clausius-Mossotti formulas

In the mid-nineteenth century, Clausis, Mossotti and Maxwell essentially gave a first order Taylor expansion for (what is now understood as) the homogenized coefficients associated with a constant background medium perturbed by diluted spherical inclusions. Such an approach was recently used and extended by the team MATHERIALS to reduce the variance in numerical approximations of the homogenized coefficients, cf. [39], [38], [62]. In [12], M. Duerinckx and A. Gloria gave the first rigorous proof of the Clausius-Mossotti formula and provided the theoretical background to analyze the methods introduced in [62].

7.2. Derivation of nonlinear elasticity from polymer-physics

7.2.1. Reconstruction of analytical constitutive laws

In [10], M. de Buhan (CNRS, Univ. Paris Descartes), A. Gloria, P. Le Tallec and M. Vidrascu proposed a numerical method to produce analytical approximations (that can be used in practical nonlinear elasticity softwares) of the numerical approximations obtained in [57] of the discrete-to-continuum energy density derived theoretically in [1]. This numerical method is based on the parametrization of the set of polyconvex Ogden laws and on the combination of a least square method and a genetic algorithm (cf. CMA-ES, https://www.lri.fr/ hansen/cmaesintro.html).

7.2.2. Stochastic homogenization of unbounded integral functionals

In [34], M. Duerinckx and A. Gloria succeeded in relaxing one of the two unphysical assumptions made in [1] on the growth of the energy of polymer chains. In particular, [34] deals with the case when the energy of the polymer chain is allowed to blow up at finite deformation.

7.3. Numerical methods

7.3.1. Numerical homogenization

Inspired by the quantitative analysis of [15] and [17], Z. Habibi (former SIMPAF post-doctoral fellow) and A. Gloria introduced in [13] a general method to reduce the so-called resonance error in numerical homogenization, both at the levels of the approximation of the homogenized coefficients and of the correctors. This method significantly extends [2]. The method relies on the introduction of a massive term in the corrector equation and of a systematic use of Richardson extrapolation. In the three academic examples of heterogeneous coefficients (periodic, quasiperiodic, and Poisson random inclusions), the method yields optimal theoretical and empirical convergence rates, and outperforms most of the other existing methods.

7.3.2. Numerical methods for evolution equations

In [11], G. Dujardin and P. Lafitte (ECP) published a result on the asymptotic behavior of splitting schemes applied to multiscale systems which have strongly attracting equilibrium states. They proposed a definition of the asymptotic order of such schemes and proved on examples of ODEs and PDEs systems that one can achieve high asymptotic order with such schemes, provided sufficient conditions are fulfilled.

In [25], G. Dujardin proposed to use high order methods for the numerical simulation of rotating Bose-Einstein condensates. With his co-authors, he developed exponential Runge-Kutta methods and Lawson method for this problem and he analyzed the convergence order of these methods. In particular, they proved that one can achieve maximal order $2s$ with methods with s-stages. They also supported their analysis with numerical experiments carried out in physically realistic simulations.

7.4. Schrödinger equations

7.4.1. Nonlinear optic fibers

In [18], S. De Bièvre, G. Dujardin, and S. Rota-Noradi, in collaboration with physicists of the PhLAM laboratory in Lille, developed an analysis of the phenomenon of modulational instability in dispersion-kicked optical fibers. They proposed a genuine analysis of the phenomenon, together with estimates on physical properties such as the gain along the fibers, and they showed that their analysis actually fits both numerical and physical experiments.

In [20], S. De Bièvre and G. Dujardin, in collaboration with physicists of the PhLAM laboratory in Lille, developed an analysis of the propagation along a periodically-modulated optic fiber of generalized Peregrine rogue waves. In particular, they provided a full analysis of the multiple compression points appearing in such waves.

In D. Bonheure and R. Nascimento [21] obtained new results on the existence and qualitative properties of waveguides for a mixed-diffusion NLS. They provided a full qualitative description of the waveguides when the fourth order dissipation is small.

7.4.2. *Nonlinear Schrödinger equations*

S. De Bièvre, S. Rota Nodari, and F. Genoud (CEMPI visitor, September 2013) have explained the geometry underlying the so-called energy-momentum method for proving orbital stability in infinite dimensional Hamiltonian systems. Applications include the orbital stability of solitons of the NLS and Manakov equations. This work appeared as a chapter (120p) in the first volume of the CEMPI Lecture Notes in Mathematics, cf. [48].

In [26], Bonheure, S. Cingolani and M. Nys obtained new striking results on stationary solutions of the 3D NLS driven by an exterior magnetic field. They construct a new class of cylindrical solutions in the energy class which concentrate, in the semi-classical limit, on a circle of the plane through the equator. In contrast with the case of solutions localized around a single point, the concentration is driven by the electrical field as well as the magnetic field

7.5. Stochastic acceleration and approach to equilibrium

S. De Bièvre and E. Soret rigorously proved the growth rate of the energy in a Markovian model for stochastic acceleration of a particle in a random medium, cf. [67] and [7].

S. De Bièvre, Carlos Mejia-Monasterio (Madrid) and Paul E. Parris (Missouri) [49] studied thermal equilibration in a two-component Lorentz gas, in which the obstacles are modeled by rotating disks. They show that a mechanism of dynamical friction leads to a fluctuation-dissipation relation that is responsible for driving the system to equilibrium.

Stephan De Bièvre, JeremyFaupin (Metz) and Schuble (Metz) [32] studied a related model quantum mechanically. Here a quantum particle moves through a field of quantized bose fields, modeling membranes that exchange energy and momentum with the particle. They establish a number of spectral properties of this model, that will be essential to study the time-asymptotic behavior of the system.

7.6. Miscellaneous results

In [24] A. Benoit proved that for linear hyperbolic systems of equations in the quarter space a violent instability can be caused by the accumulation of an arbitrary large number of weak instabilities. The proof of this result is based on the construction of the WKB expansions for hyperbolic corner problems with self-interacting phases and is a continuiation of [45].

In [9], C. Cancès, T. Gallouët, and L. Monsaingeon gave a gradient flow interpretation for incompressible immiscible two-phase flows in porous media. With C. Chainais-Hillairet, T. Gallouët characterized the pseudo-stationary state for a corrosion model in [31].

In [8], D. Bonheure, E. Moreira dos Santos, M. Ramos and H. Tavares construct least energy nodal solutions of Hamiltonian elliptic systems. The construct is tricky since the functional associated to Hamiltonian elliptic systems is strongly indefinite. The proof uses a dual variational argument and an approximation scheme with some ideas of Gama-convergence type.

In [27], D. Bonheure, P. D'Avenia and A. Pomponio aim to derive rigorously the PDE formulation of the Born-Infeld model in the electrostatic case. This nonlinear model of electromagnetism was introduced by Born and Infeld who proposed a new Lagrangian which theoretically assumes the existence of a maximal field intensity, likewise Einstein's Lagragian of special relativity opposed to Newton's Lagrangian of classical mechanics. The paper contains new results and new insights on the model. It covers several relevant particular cases but we are still far from the full understanding of the problem.

In [29] and [30], D. Bonheure and coauthors study patterns and phase transitions in a fourth order extension of the famous Allen-Cahn model. In [29], some rigidity results à la Gibbons are proved while [30] concerns qualitative properties of positive patterns with Navier boundary conditions. A conjecture related to De Giorgi's famous one concerning the one dimensionality of monotone phase transition in the classical Allen Cahn model is proposed in [29].

In [28], D. Bonheure and collaborators study multi-layer solutions of the Lin-Ni-Takagi model, which comes from the Keller-Segel model of chemotaxis in a specific case. A remarkable feature of the results is that the layers do not accumulate to the boundary of the domain but satisfy an optimal partition problem contrary to the previous type of solutions constructed for this model.

In [53], M. Duerinckx proved a new mean-field limit result for the gradient flow evolution of particle systems with pairwise Riesz interactions, in dimensions 1 and 2, in cases for which this problem was still open. The proof is based on a method introduced by Serfaty [66] in the context of the Ginzburg-Landau vortices, using regularity and stability properties of the limiting equation.

<p align="center" style="color:red">MINT Project-Team</p>

6. New Results

6.1. Physical and Perceptual Independence of Ultrasonic Vibration and Electrovibration for Friction Modulation

Eric Vezzoli, Wael Ben Messaoud, Michel Amberg, Betty Lemaire-Semail, Frédéric Giraud, Marie-Ange Bueno

Two different principles are available to modulate the user perceived roughness of a surface: electrovibration and ultrasonic vibration of a plate. The former enhances the perceived friction coefficient and the latter reduces it. In this work, we highlighted the independence of the two effects on the physical and perceptual point of view to confirm the increased range of sensation that can be supplied by the two coupled techniques. Firstly, a tribometric analysis of the induced lateral force on the finger by the two coupled effects has been achieved, then a study on the dynamic of the two effects is reported. In the end, a psychophysical experiment on the perception of the two coupled techniques confirms the approach.

6.2. Preliminary design of a multi-touch ultrasonic tactile stimulator

Sofiane Ghenna, Christophe Giraud-Audine, Frédéric Giraud, Michel Amberg, Betty Lemaire-Semail

Currently there is no solution able to provide a multitouch tactile stimulation based on friction reduction tactile devices. The main objective of this work is to achieve a control method which allows to have a differentiated tactile stimulation on two fingers simultaneously, by superimposing two vibration modes. The proof of concept has been established on a 1D beam, where the tactile stimulation could be differentiated on two selected positions. We have presented the key design rules, as well as the control method. Finally, a psychophysical evaluation has shown that users can detect the location of nodes and antinodes of vibration with an average success rate of 78%

6.3. Generalised modal analysis for closed-loop piezoelectric devices

Christophe Giraud-Audine, Frédéric Giraud, Michel Amberg, Betty Lemaire-Semail.

Stress in piezoelectric material can be controlled by imposing the electrical field. Thanks to a feedback, this electrical field can be a function of some strain related measurement so as to confer to the piezoelectric device a closed loop macroscopic behaviour.We address the modelling of such system by extending the modal decomposition methods to account for the closed loop. To do so the boundary conditions are modified to include the electrical feedback circuit, hence allowing a closed-loop modal analysis. A case study is used to illustrate the theory and to validate it. The main advantage of the method is that design issue such as coupling factor of the device and closed loop stability are simultaneously captured.

6.4. Pressure dependence of friction modulation in ultrasonic devices

Wael Ben Messaoud, Eric Vezzoli, Frédéric Giraud, Betty Lemaire-Semail

Ultrasonic vibrating devices are able to modulate the friction of a finger sliding on them. The underlying principles of the friction reduction are still unclear, and this work is carried out to investigate the influence of the ambient pressure on the friction modulation. A specific tactile stimulator has been used for this purpose and the friction between the finger sliding on the device has been recorded for an ambient pressure of 0.5 and 1 atm showing a significant difference for comparable experimental conditions. A comparison with the model proposed in literature is performed underlying that the squeeze film interaction can be present but not the only responsible of the friction modulation in this kind of devices.

Mjolnir Team

7. New Results

7.1. Introduction

The following sections summarize our main results of the year. For a complete list, see the list of publications at the end of this report.

7.2. HCI models, theories, and frameworks

Participants: Géry Casiez, Alix Goguey, Stéphane Huot.

Pointing is one of the most common and frequent action made with any interactive system whether it be a desktop computer, a mobile device or a wall-size display. Although it has been extensively studied in HCI, current pointing techniques provide no adequate way to select very small objects whose movements are fast and unpredictable, and theoretical tools –such as Fitts' law-– do not model unpredictable motion. To inform the design of appropriate selection techniques, we studied human performance (speed and accuracy) when selecting moving objects in a 2D environment with a standard mouse. We characterized selection performance as a function of the predictability of the moving targets, based on three parameters: the speed (S) of the target, the frequency (F) at which the target changes direction, and the amplitude (A) of those direction changes. Our results show that for a given speed, selection is relatively easy when A and F are both low or high, and difficult otherwise [22].

In spite of previous work showing the importance of understanding users' strategies when performing tasks, HCI researchers comparing interaction techniques remain mainly focused on performance. This can be explained to some extent by the difficulty to characterize users' strategies. To alleviate this problem, we introduced new metrics to quantify if an interaction technique introduces an object or command-oriented strategy, i.e. if users favor completing the actions on an object before moving to the next one, or in contrast, if they are reluctant to switch between commands [21]. Through a study comparing two novel interaction techniques with two from the literature, we showed that our metrics allow to replicate previous findings on users' strategies concerning the latter.

To our knowledge, there are no general design and evaluation methodologies available for the development of digital musical instruments (DMI). One reason is the large diversity of design and evaluation contexts possible in musical interaction, e.g. is this evaluation done from the perspective of the DMI designer/manufacturer, the musician playing it, or the audience watching it be performed? With our collaborators of the MIDWAY associate team, we have analyzed all papers and posters published in the proceedings of the NIME conference from 2012 to 2014 [16]. For each publication that explicitly mentioned the term "evaluation", we looked for: a) What targets and stakeholders were considered? b) What goals were set? c) What criteria were used? d) What methods were used? e) How long did the evaluation last? Results show different understandings of evaluation, with little consistency regarding the usage of the word. Surprisingly in some cases, not even basic information such as goal, criteria and methods were provided. Beyond the attempt to provide an idea of what "evaluation" means for the NIME community, we pushed the discussion towards how we could make a better use of it and what criteria should be used regarding each goal.

7.3. Transfer functions and latency

Participants: Géry Casiez, Alix Goguey, Stéphane Huot, Sylvain Malacria, Nicolas Roussel.

Our work on transfer functions mainly focused this year on *edge-scrolling*, which allows users to scroll a viewport while simultaneously dragging near or beyond its edge. Common implementations rely on rate control, mapping the distance between the pointer and the edge of the viewport to the scrolling velocity. While ubiquitous in operating systems, edge-scrolling had received little attention, though previous works suggested that rate control may be suboptimal for isotonic pointing devices (e.g. mice and touchpads) and space beyond the window's edge might be scarce, limiting scrolling control. To address these problems, we developed *Push-Edge* and *Slide-Edge* two position-based techniques that allow scrolling by "pushing" against the viewport edge [23]. A controlled experiment shows that our techniques reduce overshoots and offer performance improvements up to 13% over traditional edge-scrolling.
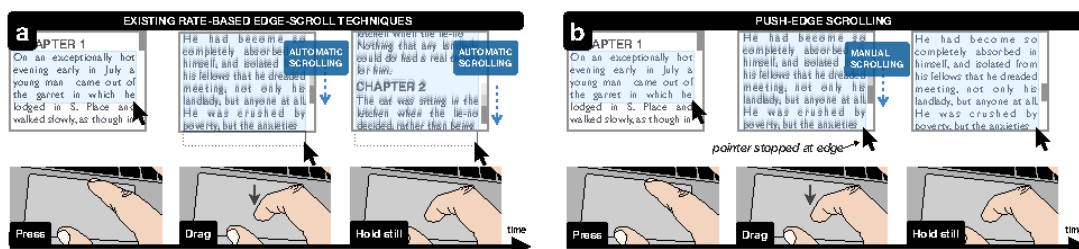

Figure 1. When selecting text with a touchpad, downward movements after crossing the viewport edge will (a) change the rate of automatic scrolling with existing techniques (rate control); or (b) manually scroll the document, stopping the pointer at the edge with push-edge scrolling (position control).

Our work on latency focused on its measurement in existing graphical user interfaces, a problem for which we developed a simple method [18], [27]. Our method consists in positioning an unmodified optical mouse on the screen while displaying and translating a particular texture to fake mouse displacements, which results in controlled mouse events. This works with most optical mice and allows accurate and real-time latency measures up to 5 times per second. The method also allows easy insertion of probes at different places in the system to investigate the sources of latency. Measurements performed on different systems, toolkits and applications notably showed that latency is affected by the operating system and system load. Substantial differences were also found between C++/GLUT and C++/Qt or Java/Swing implementations, as well as between Web browsers.

7.4. Interaction techniques

Participants: Géry Casiez, Stéphane Huot, Sylvain Malacria.

While touchpads are both widespread and expressive input devices, there has been surprisingly little research regarding how they could be used for more than simple pointer movements. In [17], we explore the design space of gesture shortcuts on touchpads and introduce four novel interaction techniques. *SpotPad* and *LociPad* rely on one-finger static gestures, but differ in their graphical representation. *ChordPad* relies on two-finger static gestures with a hierarchical representation. Finally, *InOutPad* relies on dynamic gestures crossing the edges of the touchpad. We compare the properties of these four interaction techniques and describe how they can be deployed on OS X.

The hands of virtual characters are highly complex 3D models that can be tedious and time-consuming to animate with current methods. In [14], we introduce *THING*, a novel tablet-based approach that leverages multi-touch interaction for the quick and precise control of a 3D hand's pose. The flexion/extension and abduction/adduction of the virtual fingers can be controlled individually for each finger or for several fingers in parallel through sliding motions on the tablet. We describe two variants of THING: MobileTHING, which

maps the spatial location and orientation of the tablet to that of the virtual hand, and DesktopTHING, which combines multi-touch controls of fingers with traditional mouse controls for the hand's global position and orientation. We also report on two usability studies in which we compared THING to mouse-only controls and a data glove.

Figure 2. THING enables the control of 3D hand models (in blue) by sliding fingers along sliders arranged in a morphologically-consistent pattern on the tablet's screen.

Interactive technologies have radically changed the way visual artists work, and a large portion of the artistic production has now moved from paper to the computer. However, many artists still work on paper and keep using traditional painting and drawing tools. This is not only due to resistance to progress or due to the well-known usability properties of physical tools: despite the use of pen displays, the progress of artistic stroke-rendering techniques, and the powerful and advanced functionalities of existing computer tools, they fail to fully capture the richness and variety of artistic styles supported by physical media. We interviewed four professional illustrators in their work environment. We also followed the work of an artist for a two-year period. We observed that artists mix a variety of techniques that involve specialized computer software and hardware such as Adobe Photoshop, a graphics tablet and a scanner, and traditional physical tools such as pencils, paper, and customized light tables. Our findings inspired *BricoSketch* [25], an augmented paper interface that enables illustrators to zoom into parts of their drawings and work at different levels of detail on paper. Our early results show that BricoSketch supports real tasks, improving productivity on paper while enhancing illustrators' creative ways of working.
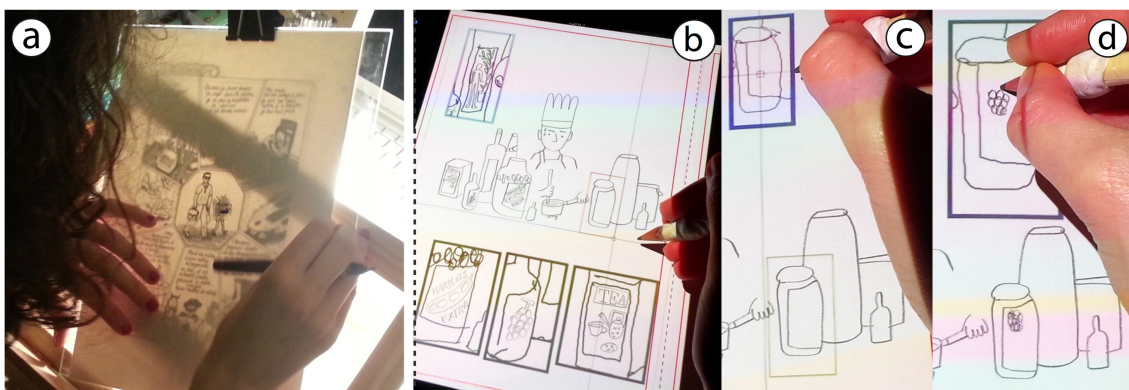

Figure 3. BricoSketch. (a) An artist works on layers of physical paper with a home-made light table: she draws the panels of a page for a graphic novel by using earlier sketches as guides. (b-d) The same artist uses our system to add details to different parts of her illustration through partial scaled views.

7.5. Interactive visualization

Participant: Fanny Chevalier.

The differential diagnosis of hereditary disorders is a challenging task for clinicians due to the heterogeneity of phenotypes that can be observed in patients. Existing clinical tools are often text-based and do not emphasize consistency, completeness, or granularity of phenotype reporting, which can impede clinical diagnosis and limit their utility to genetics researchers. The *PhenoBlocks* tool described in [13] is a novel visual analytics platform designed to support clinical differential diagnosis. It supports the comparison of phenotypes between patients, or between a patient and the hallmark features of a disorder. An informal evaluation with expert clinicians suggests that the visualization effectively guides the process of differential diagnosis and could reinforce the importance of complete, granular phenotypic reporting.
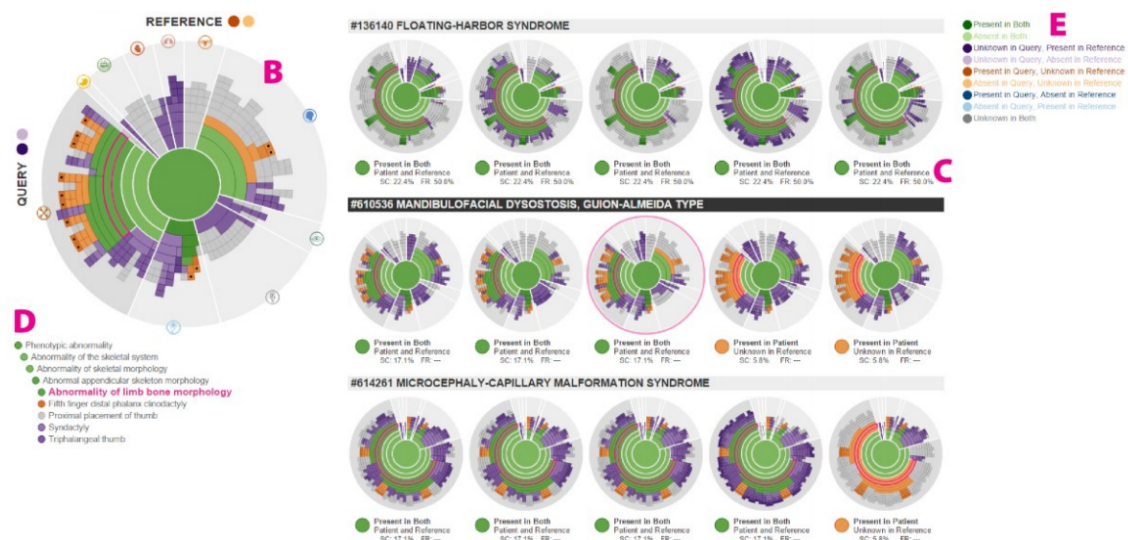

Figure 4. PhenoBlocks allows to compare the phenotype hierarchies of an undiagnosed query patient to a diagnosed reference patient. During differential diagnosis, clinicians use shared (green) phenotypes to gauge confidence in their diagnostic hypothesis and missing (purple) phenotypes to identify candidates for subsequent analysis.

7.6. Adaptive interfaces

Participant: Sylvain Malacria.

As news is increasingly accessed on smartphones and tablets, the need for personalizing news applications is apparent. In [19], [20], we report on a series of studies addressing key issues in the development of adaptive news interfaces. We first surveyed users' news reading preferences and behaviors. We then implemented and deployed an Android application that logs users' interactions with the application. We used the logs to train a classifier and showed that it is able to reliably recognize a user according to their reader type. Finally we evaluated alternative, adaptive user interfaces for each reader type. The evaluation demonstrates the differential benefit of the adaptation for different users and the feasibility of adaptive interfaces for news applications.

In [12], we investigate the use of a companion application on a tablet to augment viewing of information-rich television programs. The application displays a synchronized graphical abstraction of the program's content in the form of a concept map. Two experiments were conducted involving participants watching an astronomy documentary. Results show that the companion application improved participants' understanding and recall of the program. Participants were found to manage their visual attention systematically when using the companion application, and correlations were found in the way they shifted their gaze from TV screen to tablet and back in response to changes in the program content. Increasing interaction with the application disrupted understanding of the television program and visual attention. Participants were positive about the value of companion applications for understanding and recall of programs, but distraction and "knowing where to look" were significant concerns.

MODAL Project-Team

7. New Results

7.1. Functional data analysis applied to hydrological or environnemental data

Participant: Sophie Dabo.

The new results concern particularly functional data analysis applied to hydrological or environmental data. First in a recent paper ([16]), two statistical techniques from the theory of functional data classification are adapted and applied for the analysis of flood hydrographs. Functional classification directly employs all data of a discharge time series and thus contains all available information on shape, peak, and timing. This potentially allows a better understanding and treatment of floods as well as other hydrological phenomena.

7.2. New functional regression model when data are auto-correlated

Participant: Sophie Dabo.

We develop a new functional regression model when data are auto-correlated, in collaboration with Serge Guillas (University of College London) and Camille Ternynck (University of Lille 2). This work will appear in Journal of Multivariate Analysis. (Dabo-Niang, S, Guillas, S et Ternynck, C. (2016). More efficient kernel functional spatial regression estimation with autocorrelated errors. *Journal of Multivariate Analysis*). In this work we introduce a new procedure for the estimation in the nonlinear functional regression model where the explanatory variable takes values in an abstract function space and the residual process is autocorrelated. The procedure consists in a pre-whitening transformation of the dependent variable based on the estimated autocorrelation. We establish both consistency and asymptotic normality of the regression function estimate. For kernel methods encountered in the literature, the correlation structure is commonly ignored (the so-called "working independence estimator"); we show here that there is a strong benefit in taking into account the autocorrelation in the error process. We also find that the improvement in efficiency can be large in our functional setting, up to 25% in the presence of high autocorrelation levels. Concerning spatial data, we develop a new spatial prediction method that takes into account the spatial dependence. This work will appear in Journal of Nonparametric Statistics (Dabo-Niang, Ternynck, C., Yao, A.-F. (2016). Nonparametric prediction in the multivariate spatial context. *Journal of Nonparametric Statistics*)

7.3. Differential gene expression analysis

Participant: Guillemette Marot.

The use of empirical Bayesian techniques implemented in the Bioconductor package `limma` has enabled to better understand Waldenstrom's macroglobulinemia. Gene Set enrichment analysis was also performed after differential analysis. The new findings in Biology have been published in [21].

7.4. Evolutionary clustering for categorical data

Participant: Julien Jacques.

This is a joint work with Md Abul Hasnat, Julien Velcin and Stephane Bonnevay (Univ. de Lyon).

An evolutionary clustering algorithm for categorical data has been developed, based on parametric links between multinomial mixture models. This model has been used to study the evolution of opinions in Twitter data. A Preprint of this work is available [54].

7.5. Clustering categorical functional data: Application to medical discharge letters

Participants: Cristian Preda, Cristina Preda, Vincent Vandewalle.

Categorical functional data represented by paths of a stochastic jump process are considered for clustering. For paths of the same length, the extension of the multiple correspondence analysis allows the use of well-known methods for clustering finite dimensional data. When the paths are of different lengths, the analysis is more complex. In this case, for Markov models we have proposed an EM algorithm to estimate a mixture of Markov processes. This work has been presented in a conference [34].

7.6. Degeneracy in Gaussian Mixtures with missing data

Participants: Christophe Biernacki, Vincent Vandewalle.

The missing data problem is well-known for statisticians but its frequency increases with the growing size of modern datasets. In Gaussian model-based clustering, the EM algorithm easily takes into account such data by dealing with two kinds of latent levels: the components and the variables. However, the quite familiar degeneracy problem in Gaussian mixtures is aggravated during the EM runs. Indeed, numerical experiments clearly reveal that degeneracy is quite slow and also more frequent than with complete data. In practice, such situations are difficult to detect efficiently. Consequently, degenerated solutions may be confused with valuable solutions and, in addition, computing time may be wasted through wrong runs. A simple condition on the latent partition to avoid degeneracy has been exhibited, and a constrained version of the Stochastic EM (SEM) algorithm satisfying this condition has been proposed. This work has been presented in a conference [33].

7.7. Model for conditionally correlated categorical data

Participants: Christophe Biernacki, Vincent Vandewalle, Matthieu Marbac-Lourdelle.

An extension of the latent class model is proposed for clustering categorical data by relaxing the classical class conditional independence assumption of variables. In this model (called CCM for Conditional Correlated Model), variables are grouped into inter-independent and intra-dependent blocks in order to consider the main intra-class correlations. The dependence between variables grouped into the same block is taken into account by mixing two extreme distributions, which are respectively the independence and the maximum dependence ones. In the conditionally correlated data case, this approach is expected to reduce biases involved by the latent class model and to produce a meaningful model with few additional parameters. The parameters estimation by maximum likelihood is performed by an EM algorithm while a MCMC algorithm avoiding combinatorial problems involved by the block structure search is used for model selection. Applications on sociological and biological data sets bring out the proposed model interest. These results strengthen the idea that the proposed model is meaningful and that biases induced by the conditional independence assumption of the latent class model are reduced. This work is published [20] . Furthermore, an R package (Clustericat) is available on Rforge(see https://github.com/rforge/clustericat).

7.8. Model-based clustering for multivariate partial ordinal data

Participants: Christophe Biernacki, Julien Jacques.

We design the first univariate probability distribution for ordinal data which strictly respects the ordinal nature of data. More precisely, it relies only on order comparisons between modalities, the proposed distribution being obtained by modeling the data generating process which is assumed, from optimality arguments, to be a stochastic binary search algorithm in a sorted table. The resulting distribution is natively governed by two meaningful parameters (position and precision) and has very appealing properties: decrease around the mode, shape tuning from uniformity to a Dirac, identifiability. Moreover, it is easily estimated by an EM algorithm since the path in the stochastic binary search algorithm is missing. Using then the classical latent class assumption, the previous univariate ordinal model is straightforwardly extended to model-based clustering for multivariate ordinal data. Again, parameters of this mixture model are estimated by an EM algorithm. Both simulated and real data sets illustrate the great potential of this model by its ability to parsimoniously identify particularly relevant clusters which were unsuspected by some traditional competitors. This work is now published in an international journal [12] and is also currently available in the MixtComp software at https://modal-research.lille.inria.fr/BigStat/

7.9. Semi-Linear Auto-Associative Model

Participant: Serge Iovleff.

We design a new model for data analysis which is a generalization of the probabilistic PCA. The interpretation properties of the PCA are preserved while presence of non-linear repartitions in data can be detected and adjusted using B-spline regression. This model has been published in [18].

6. New Results

6.1. Homogeneity Theory

Homogeneity is one of the tools we develop for finite-time convergence anasysis. In 2015 this concept has received various improvements:

- The concept of homogeneous evolution equation in a Banach space has been introduced in [67]. It provides the background for the extencion of all homogeneity-based tools for control design and analysis to distributed parameters systems.

- Scalability is a property describing the change of the trajectory of a dynamical system under a scaling of the input stimulus and of the initial conditions. Particular cases of scalability include the scale invariance and fold change detection (when the scaling of the input does not influence the system output). In the paper [19] is shown that homogeneous systems have this scalability property while locally homogeneous systems approximately possess this property.

- In the paper [25] the notion of homogeneity in the bi-limit is extended to local homogeneity and then to homogeneity in the multi-limit. The converse Lyapunov/Chetaev theorems on (homogeneous) system instability are obtained. The problem of oscillation detection for nonlinear systems is addressed. The sufficient conditions of oscillation existence for systems homogeneous in the multi-limit are formulated.

- The notion of weighted homogeneity is extended in [81] to the time-delay systems. It is shown that the stability/instability of homogeneous functional systems on a sphere implies the global stability/instability of the system. The notion of local homogeneity is introduced, a relation between stability/instability of the locally approximating dynamics and the original time-delay system is established using Lyapunov-Razumikhin approach

- In [27] global delay independent stability is analyzed for nonlinear time-delay systems by applying homogeneity theory. It is shown that finite-time stability can be encountered in this class of systems under uniformity of the convergence time with respect to delay. Some additional tools for stability analysis of time-delay systems using homogeneity are also presented: in particular, it is shown that if a time-delay system is homogeneous with nonzero degree and it is globally asymptotically stable for some delay, then this property is preserved for any delay value, which is known as the independent of delay (IOD) stability.

- Theorems on Implicit Lyapunov Functions for finite-time and fixed-time stability analysis of nonlinear systems are presented in [37]. Based on these resutls, new homogenenous nonlinear control laws are designed for robust stabilization of a chain of integrators. The presented results are extended to Multi-Input Multi-Output in [38]. A time-suboptimal control design algorithm based Implicit Lyapunov Function Method is developed in [40]. A robustness-oriented comparison of the optimal and suboptimal solutions in practical implementations of the proposed controller is performed via the numerical example of double integrator. A novel scheme of practical implimentation of the implicit Lypunov function-based control is developed in [79]. It replaces the implicitely defined Lyapunov function (in the feedback law) with the homogenenous norm of the state. Such a modification simplifies the practical application of the finite-time stabiliting feedback control.

- The uniform stability notion for a class of nonlinear time-varying systems is studied in [42] using the homogeneity framework. It is assumed that the system is weighted homogeneous considering the time variable as a constant parameter, then several conditions of uniform stability for such a class of systems are formulated. The results are applied to the problem of adaptive estimation for a linear system. The detailed report on time-varying homogenety is given in [83].

- In the paper [52] we consider the continuous homogeneous observer defined in the case of the triple integrator. Originally, convergence of the algorithm was only proved when the degree of homogeneity was sufficiently close to 0 without more tractable information. We show here that, in the case of the triple integrator, the observer presents global finite-time stability for any negative degree under constructive conditions on the gains. This is achieved with a homogeneous Lyapunov function design.

- The work [61] addresses the stabilization of dynamical systems in presence of uncertain bounded perturbations using theory. Under some assumptions, the problem is reduced to the stabilization of a chain of integrators subject to a perturbation and is treated in two steps. The evaluation of the disturbance and its compensation. Homogeneous observer and control are the tools utilized to achieve a global asymptotic stability and robustness. The result is formally proven and, to validate the theory, it is applied to the control of the telescopic link of a hydraulic actuated industrial crane used in forestry.

- A geometric homogeneity of evolution equation in a Banach space is introduced in [67]. Scalability property of solutions of homogeneous evolution equations is proven. Some qualitative characteristics of stability of trivial solution are also provided. In particular, finite-time stability of homogeneous evolution equations is studied. Theoretical results are supported by examples from mathematical physics.

- The second order planar nonlinear affine control problem is studied [69]. A homogeneous robust finite-time stabilizing control is developed for the most general case of matched and, the more challenging, mismatched nonlinear perturbations. A homogeneous observer is designed for the planar system. Explicit restrictions on the observer gains and nonlinearities are presented. The main contribution lies in the proposed combination of the explicit and implicit Lyapunov function methods as well as weighted homogeneity while providing finite-time stability analysis.

6.2. Algebraic Technique For Estimation, Differentiation And Its Applications

Algebraic technique is the other tool we develop for providing finite-time convergence.

- The integer order differentiation by integration method based on the Jacobi orthogonal polynomials for noisy signals was originally introduced by Mboup, Join and Fliess. The paper [35] proposes an extention of this method from the integer order to the fractional order to estimate the fractional order derivatives of noisy signals. Two fractional order differentiators are deduced from the Jacobi orthogonal polynomial filter, using the Riemann-Liouville and the Caputo fractional order derivative definitions respectively. Exact and simple formulas for these differentiators are given by integral expressions. Some error bounds are provided for the corresponding estimation errors. The noise error contribution due to a large class of stochastic processes is studied in discrete case.

- Armed with structures, group sparsity can be exploited to improve the performance of adaptive estimation. In the paper [45], the adaptive estimation algorithm for cluster structured sparse signals, called A-CluSS, is proposed. In particular, a hierarchical Bayesian model is built, where both sparse prior and cluster structured prior are exploited simultaneously. The adaptive updating formulas for statistical variables are obtained via the variational Bayesian inference and the resulted algorithms can adaptively estimate the cluster structured sparse signals without knowledge of block size, block numbers and block locations. In [75], a group sparse regularized least-mean-square (LMS) algorithm is proposed to cope with the identification problems for multiple/multi-channel systems. An iterative online algorithm is proposed via proximal splitting method.

6.3. Set-Theoretic Methods of Control And Estimation

Interval and ellipsoidal estimations can be regarded as particular finite-time algorithms, since they provide guaranteed estimates of the values from the intial time. We develop these tools for some years now.

- An approach to interval observer design for Linear Parameter-Varying (LPV) systems is proposed in [20]. Stability conditions are expressed in terms of matrix inequalities. Applying L1/L2 framework the robustness and estimation accuracy with respect to model uncertainty are analyzed.

- New delay-dependent conditions of positivity for linear systems with time-varying delays are introduced in [56]. These conditions are applied to interval observer design for systems with time-varying delays in the state equations and in the measurements. In [28] the problem of interval observer design is addressed for a class of descriptor linear systems with time delays. An interval observation for any input in the system is provided. The control input is designed together with the observer gains in order to guarantee interval estimation and stabilization simultaneously. Efficiency of the proposed approach is illustrated by numerical experiments with Leontief delayed model.

- The work [29] is devoted to interval observers design for discrete-time Linear Parameter-Varying (LPV) systems under the assumption that the vector of scheduling parameters is not available for measurements. Two problems are considered: a pure estimation problem and an output stabilizing feedback design problem where the stability conditions are expressed in terms of Linear Matrix Inequalities (LMIs).

- The paper [48] investigates the interval observer design for a class of nonlinear continuous systems, which can be represented as a superposition of a uniformly observable nominal subsystem with a Lipschitz nonlinear perturbation. It is shown in this case there exists an interval observer for the system that estimates the set of admissible values for the state consistent with the output measurements. In [77] similar methodology is extended to singular systems.

- A finite-time version, based on Implicit Lyapunov Functions, for the Attractive Ellipsoid Method is developed in [65]. Based on this, a robust control scheme [36] is presented to ensure finite-time convergence of the solutions of a chain of integrators with bounded output perturbations to a minimal ellipsoidal set. The control parameters are obtained by solving a minimization problem of the " size " of the ellipsoid subject to a set of Linear Matrix Inequalities, and by applying the implicit function theorem.

- In [78] we consider a problem of sliding mode control design for LTI systems with multiplicative disturbances of the input and noisy measurements of the output. We apply the minimax observer to provide the best possible estimate of the system's state. Then we solve a problem of optimal reaching for the observer: we design sub-optimal control algorithms generating continuous and discontinuous feedback controls that steer the observer as close as possible to a given sliding hyperplane in a finite time.

6.4. Observability And Observer Design For Nonlinear Systems

- In [18] a method to carried out the state estimation is proposed for a class of nonlinear systems with unknown inputs whose dynamics is governed by differential-algebraic equations (DAE). We achieve, under suitable conditions, to replace the original DAE for a system with differential equations only by using a zeroing manifold algorithm inducing a state space dimension reduction.

- In the paper [44], we investigate the estimation problem for a class of partially observable nonlinear systems. For the proposed Partial Observer Normal Form (PONF), necessary and sufficient conditions are deduced to guarantee the existence of a change of coordinates which can transform the studied system into the proposed PONF.

- Using the theory of non-commutative rings, the delay identification problem of nonlinear time-delay systems with unknown inputs is studied in the paper [82]. Necessary and sufficient conditions are proposed to judge the identifiability of the delay, where two different cases are discussed for the dependent and independent outputs, respectively. After that, necessary and sufficient conditions are given to analyze the causal and non-causal observability for nonlinear time-delay systems with unknown inputs.

- In the paper [58], we investigate the stabilization of a linear plant subject to network constraints, partial state knowledge and time varying bounded parameter uncertainties. An event–triggered version of the Luenberger observer is proposed, and necessary conditions on the uncertainties are given in term of LMI's to enable output– based stabilization under different triggering strategies.

- The papers [47], [76] investigate an unknown input observer design for a large class of linear systems with unknown inputs and commensurate delays. A Luenberger-like observer is proposed by involving only the past and actual values of the system output. The required conditions for the proposed observer are considerably relaxed in the sense that they coincide with the necessary and sufficient conditions for the unknown input observer design of linear systems without delays.

- The paper [71] deals with the problem to estimate some states of a multi-output nonlinear dynamical system which is partially observable. To address this problem, this paper provides a set of geometrical conditions that guarantee the existence of a change of coordinates which decomposes the studied nonlinear dynamical system into two dynamical subsystems, where the first one is of the well-known output injection form. This transformed form allows us to design a simple reduced-order (Luenberger-like) observer to estimate the observable state.

6.5. Model-Free Control

- In the paper [41] the Universal Integral Control, introduced by H.K. Khalil, is revisited by employing mollifiers instead of a high-gain observer for the differentiation of the output signal. The closed loop system is a classical functional differential equation with distributed delays on which standard Lyapunov arguments are applied to study the stability. Low-pass filtering capability of mollifiers is demonstrated for a high amplitude and rapidly oscillating noise.

- The paper [64] proposes an universal adaptive control structure for robot manipulators, without knowing the dynamic model of the system, as well it is robust to corrupt payload change and initial conditions.

- In [66], the control design of an artificial pancreas, a hot research topic in diabetology, is tackled via the newly introduced model-free control and its corresponding "intelligent" proportional controller, which were already quite successful in many concrete and diverse situations. It results in an insulin injection for type 1 diabetes which displays via constant references a good nocturnal/fasting response, but unfortunately a poor postprandial behavior due to long hyperglycemia. When a variable reference is introduced, that switches between a constant one, when glycemia is more or less normal or moderate, and an exponential decay reference path, when a high glycemia rate indicates a meal intake, the results in silico, which employ real clinical data, become excellent. We obtain a bolus-shaped insulin injection rate during postprandial phases. The hyperglycemic peaks are therefore lowered a lot.

6.6. Sliding Mode Control And Estimation

- The paper [22] addresses the problem of oscillatory failure case detection in the electrical flight control system of a generic commercial airplane. A non-homogeneous differentiator is first used to provide accurate derivatives in noisy environment and fast convergence time. In this study case, fault detection is addressed in the unknown input estimation issue for fault reconstruction with the same evaluation techniques currently employed in Airbus A380 airplanes. Performance and robustness of the developed monitoring strategy are assessed using a high-fidelity Airbus benchmark and a parametric test campaign for the flight scenarios defined in the EU-FP7 ADDSAFE project.

- A new sliding mode control approach is introduced in [26] with the dedicated mathematical tools. A time-delay modification/approximation of sign function is proposed in [57], and it is shown that by substituting this new "sign " realization in the conventional sliding mode algorithms the main advantages of the sliding mode tools are preserved (like rejection of matched disturbances and hyperexponential convergence), while the chattering is reduced.

- The article [34] proposes a convex optimization approach for the design of relay feedback controllers. The case of linear systems is studied in the presence of matched perturbations. The system input is a generalized relay that may take values in a finite set of constant vectors. A simple design method is proposed using Linear Matrix Inequalities (LMIs).

- In the note [53] we study the effect of an implicit Euler time-discretization method on the stability of the discretization of a globally fixed-time stable, scalar differential inclusion representing a simple nonlinear system with a set-valued signum controller. The controller nonlinearity is a cubic term and it is shown that the fully-implicit method preserves the global Lyapunov stability property of the continuous-time system, contrarily the explicit discretization which does not. It allows to obtain finite-time convergence to the origin when the plant is undisturbed, while the cubic term provides the hyper-exponential convergence rate.

- The problem of finite-time stabilization of multi-input linear system by means of sliding mode relay feedback is considered in [68]. A new control design procedure, which combines convex embedding technique with implicit Lyapunov function method, is developed. The issues of practical implementation of the obtained implicit relay feedback are discussed. Theoretical result is supported by numerical simulation.

6.7. Non-Linear, Sampled-Data And Time-Delay Systems

- The method of Implicit Lyapunov-Krasovski Functional (ILKF) for stability analysis of time-delay systems is introduced in [39]. Theorems on Lyapunov, asymptotic, (hyper) exponential , finite-time and fixed-time stability analysis using ILKF are presented. The hyper exponential stabilization algorithm for a time-delay system is presented.

- A recent generalization of the classical ISS theory to multistable systems is presented in [17]. Based on it a robust synchronization protocols with respect to a compact invariant set of the unperturbed system are designed in [14], [49].

- The paper [21] deals with the design of an active fault-tolerant control strategy based on the supervisory control approach technique for linear time invariant MIMO systems affected by disturbances, measurement noise, and faults. From a bank of Luenberger observers that plays the role of a fault detection and isolation scheme, the supervisory algorithm selects the suitable fault-tolerant controller by means of a hysteresis-based switching mechanism based on the method proposed in this paper.

- In [30] motivated by the problem of phase-locking in droop-controlled inverter-based microgrids with delays, the recently developed theory of input-to-state stability (ISS) for multistable systems is extended to the case of multistable systems with delayed dynamics. Sufficient conditions for ISS of delayed systems are presented using Lyapunov-Razumikhin functions. It is shown that ISS multistable systems are robust with respect to delays in a feedback [55].

- The work [31] aims at enlarging the sampling intervals in several state feedback control situations by designing a sampling map in the state space. For linear time invariant (LTI) systems with state-bounded perturbations this guarantees exponential stability with a chosen decay-rate. The approach is based on linear matrix inequalities (LMIs) obtained thanks to Lyapunov-Razumikhin stability conditions and convexification arguments. Then, the obtained results are extended to design the sampling map in three dynamic sampling control situations: event-triggered control, self-triggered control, and state-dependent sampling.

- In the paper [43] the problem of discrete and continuous state estimation for a class of uncertain switched LPV systems is addressed. Parameter identification techniques are applied to realize an approximate identification of the scheduled parameters of a switched LPV system with certain uncertainties and/or disturbances. A discrete state estimation is achieved using the parameter identification. A Luenberger-like hybrid observer, based on discrete state information and LMIs approach, is used for the continuous state estimation.

- The paper [70] contributes to the exponential stability analysis for impulsive dynamical systems based on a vector Lyapunov function and its divergence operator. The method relies on a 2D time domain representation. The results are applied to analyze the exponential stability of linear impulsive systems based on LMIs.

6.8. Networked Systems

- The problem of phase synchronization for a population of genetic oscillators (circadian clocks, synthetic oscillators, etc.) is considered in the paper [13]. The proposed analysis approach is based on the Phase Response Curve model of an oscillator. The performance of the obtained solutions is demonstrated via computer experiments for two different models of circadian/genetic oscillators.

- The paper [23] focuses on the design of fixed-time consensus for first order multi-agent systems with unknown inherent nonlinear dynamics. A distributed control protocol, based on local information, is proposed to ensure the convergence of the tracking errors in finite time. Some conditions are derived to select the controller gains in order to obtain a prescribed convergence time regardless of the initial conditions.

- The problem of phase regulation for a population of oscillating systems is considered in [24] based on a Phase Response Curve (PRC) model of an oscillator. The problem of phase resetting for a network of oscillators is solved by applying a common control input. Performance of the obtained solutions is demonstrated via computer simulation for three different models of circadian/neural oscillators.

6.9. Applications

- The problem of avoiding obstacles while navigating within an environment for a Unicycle-like Wheeled Mobile Robot (WMR) is of prime importance in robotics. The work [32] solves such a problem proposing a perturbed version of the standard kinematic model able to compensate for the neglected dynamics of the robot. The effectiveness of the solution is proved, supported by experiments and finally compared with the Dynamic Window Approach (DWA) to show how the proposed method can perform better than standard methods. The paper [60] presents a decentralized solution to control a leader-follower formation of unicycle wheeled mobile robots allowing collision and obstacle avoidance. The work [62] solves the obstacle avoidance problem extending the Potential Field (PF) method for a mobile robot. The usual definition of the PF has been modified to have a field which is continuous everywhere. It is shown that the system has an attracting equilibrium at the target point, repelling equilibriums in the centers of the obstacles and saddle points on the borders. Those unstable equilibriums are avoided capitalizing on the established Input-to-State Stability (ISS) property of this multi-stable system. To escape a local minima this work makes the most of ISS property that is not lost for perturbations. And for small properly designed disturbances the global attractivity of the target point is proved.

- The paper [63] investigates the behavior of central Jacobi differentiator in robot identification applications. It is applied to compute acceleration from noisy position measurements. Its frequency domain property is analyzed via a finite impulse response (FIR) filter point of view, indicating clearly the differentiators performance. Two revolute joints planar robot parameter identification is done. Comparisons between the Jacobi differentiator and the Euler differentiation combined with Butterworth filter are drawn.

- In [50] the velocity of valve movement activity is estimated using three different differentiation schemes: an algebraic-based differentiator method, a non-homogeneous higher order sliding mode differentiator and a homogeneous finite-time differentiator. We demonstrate that this estimated velocity can be used for water quality monitoring as the differentiators can detect very rapid change in valve movements of the oyster population resulting from some external stimulus or common input.

- In the paper [15] the measurements of valve activity in a population of bivalves under natural environmental conditions (16 oysters in the Bay of Arcachon, France) are used for a physiological model identification. A nonlinear auto-regressive exogenous (NARX) model is designed and tested. Through this study, it is demonstrated that the developed dynamical model of the oyster valve movement can be used for estimating normal physiological rhythms of permanently immersed oysters and can be considered for detecting perturbations of these rhythms due to changes in the water quality, i.e. for ecological monitoring.

- Spawning observations are important in aquaculture and biological studies, and until now, such a detection is done through visual analysis by an expert. Using measurements of valve activity (i.e. the distance between the two valves) in populations of bivalves under natural environmental condition (16 oysters in the Bay of Arcachon, France, in 2007, 2013 and 2014), algorithms for an automatic detection of the spawning period of oysters are proposed in the paper [16], [51]. The fault detection method presented in the paper can also be used to detect complex oscillatory behavior which is of interest to control engineering community.

- The work presented in the paper [33] is undertaken within the European FP7 funded Advanced Fault Diagnosis for Sustainable Flight Guidance and Control (ADDSAFE) project. It proposes new fault detection and fault diagnosis techniques that could significantly help developing environmentally-friendlier aircraft. LPV model-based fault detection schemes are proposed and compared for robust and early detection of faults in aircraft control surfaces servo-loop. The proposed methodologies are based on a slight modification of the H_∞/H_- LPV optimization techniques for systems modelled in, first polytopic manner, second linear fractional representation fashion. It is shown that the proposed fault detection schemes can be embedded within the structure of in-service monitoring systems as a part of the Flight Control Computer software. Several important examples on model and signal based fault detection in aircraft Electrical Flight Control System are studied in [80].

- For analyzing the transients of induction heating systems, time-dependent phasor transformations were proposed so far in the literature. Applying these transformations to a linear R, L, C circuit equations leads to differential equations in the complex domain from which equivalent circuits modeling the envelopes of sinusoidal waveforms were derived. The work [46] proposes a phasor transformation which is based on fictitiously replacing the real voltage and current signals of a system by complex ones. It leads to transformed system equations in the real domain where instantaneous amplitudes, phases and frequencies appear explicitly, which makes the transformed equations suitable for the feedback control design. The methodology is applied to a parallel induction heating system in order to design a sliding mode controller.

- The problem of air-to-fuel ratio regulation for a direct injection engine is addressed in [54]. A LPV model of the engine is used, for which an interval observer is designed. The interval observer is applied for the model validation and control synthesis. The results of design are confirmed by implementation.

- Modular Robot Manipulators are user-configurable manipulators which provide rapid design and inexpensive implementation. To be easy-use, smart actuators embedded with position input and position feedback controller are adopted, these local controllers render the manipulators position controlled, but also result in limited performance and precision. The paper [72] targets the case that the built-in controller does not provide desirable precision for set-point regulation. Firstly a joint-level model is established, of which the nominal model can be identified with derivative observer based on the position feedback, then an auxiliary adaptive controller coping with parametric uncertainty is proposed which leads to an error close to zero, a switching control strategy is introduced considering the actuator saturation. The paper [73] addresses the set-point control of actuators integrated with built-in controller, which presents steady-state error (SSE) under certain load. To eliminate the SSE, a model of the actuator-plus-controller system is established and identified, a switched adaptive controller is developed to work with the embedded one, considering the physical constraints, a switching control strategy is proposed. The proposed algorithms are implemented on a 5-DOF modular manipulator, with comparison to classic integral controller.

- The communication [74] is devoted to a comparison between various meteorological forecasts, for the purpose of energy management, via different time series techniques. The first group of methods necessitates a large number of historical data. The second one does not and is much easier to implement, although its performances are today only slightly inferior. Theoretical justifications are related to methods stemming from a new approach to time series, artificial neural networks, computational intelligence and machine learning.

- ALINEA is a well known ramp metering closed-loop control the aim of which is to improve highway traffic. The report [84] shows that ALINEA may be slightly modified in order to be efficiently implemented without any need of crucial time-varying quantities, like the critical density and the free-flow speed, which are most difficult to estimate correctly online.

- For malaria patients, a usual observation problem consists in estimation of sequestered parasites Plasmodium falciparum from measurements of circulating ones. The model of an infected patient is rather uncertain, and for all rates (death, transition, recruitment and infection) in the model it is assumed that only intervals of admissible values are given. In addition, the measurements of the concentration of circulating parasites are subjected by a bounded noise, while some parameters, like the rate of infection of blood cells by merozoites, are completely unknown and highly time-varying. In order to evaluate the concentration of sequestered parasites, an interval observer is designed in [85], which provides intervals of admissible value for that concentration, with the interval width proportional to the model uncertainty.

RAPSODI Team

7. New Results

7.1. Design and analysis of advanced finite volumes schemes

The fact that a numerical method is able to handle nonlinear test functions in its numerical analysis is crucial in order to ensure its physical relevance, and consequently its good behavior.

In [15], C. Cancès and C. Guichard proposed a first nonlinear numerical method to solve possibly degenerate parabolic equations with anisotropy on general simplicial meshes. The nonlinear control volume finite element (CVFE) scheme is based on P1 finite elements with mass-lumping combined with a tricky upwinding of the mobilities. The method has the remarkable property of preserving the positivity of the solutions. Moreover, it ensures the decay of the physical entropy. Its convergence is proved in [15] and numerical results are exhibited. In particular, they show that the method is first order accurate in space in standard situations, but can lack robustness w.r.t. the anisotropy in some particularly unfavorable situations.

This drawback was corrected by C. Cancès and C. Guichard in [35], where a second order in space method based on the so-called VAG scheme [57] was proposed. This method is able to handle very general grids, heterogeneous data and strong anisotropy ratios. Moreover, it preserves at the discrete level the variational structure of the continuous problem, yielding the nonlinear stability of the scheme. A complete convergence analysis was performed in [35]. The numerical results presented in [35] show that the robustness default of the first nonlinear method [15] has been corrected.

In [36], C. Cancès *et al.* proposed and analyzed a nonlinear CVFE scheme for a degenerate Keller-Segel model with anisotropic and heterogeneous diffusion tensors. The scheme is based on the one proposed in [15]. The convergence of the scheme is proved under very general assumptions. Finally, some numerical experiments are carried out to prove the ability of the scheme to tackle degenerate anisotropic and heterogeneous diffusion problems over general meshes without jeopardizing the positivity of the solutions.

In [17], C. Chainais-Hillairet, A. Jüngel and S. Schuchnigg prove the time decay of fully discrete finite-volume approximations of porous-medium and fast-diffusion equations with Neumann or periodic boundary conditions in the entropy sense. The algebraic or exponential decay rates are computed explicitly. In particular, the numerical scheme dissipates all zeroth-order entropies which are dissipated by the continuous equation. The proofs are based on novel continuous and discrete generalized Beckner inequalities.

In [18], C. Chainais-Hillairet, A. Jüngel and P. Shpartko propose and analyze a numerical scheme for a spinorial matrix-diffusion model for semiconductors. The model consists of strongly coupled parabolic equations for the electron density matrix or, alternatively, of weakly coupled equations for the charge and spin-vector densities, coupled to the Poisson equation for the electric potential. The main features of the numerical scheme are the preservation of nonnegativity and L^∞ bounds of the densities and the dissipation of the discrete free energy. The existence of a bounded discrete solution and the monotonicity of the discrete free energy are proved. The fundamental ideas are reformulations using spin-up and spin-down densities and certain projections of the spin-vector density, free energy estimates, and a discrete Moser iteration. Furthermore, numerical simulations of a simple ferromagnetic-layer field-effect transistor in two space dimensions are presented.

In [32], M. Bessemoulin-Chatard and C. Chainais-Hillairet study the large–time behavior of a numerical scheme discretizing drift–diffusion systems for semiconductors. The numerical method is finite volume in space, implicit in time, and the numerical fluxes are a generalization of the classical Scharfetter–Gummel scheme which allows to consider both linear or nonlinear pressure laws. They study the convergence of approximate solutions towards an approximation of the thermal equilibrium state as time tends to infinity, and obtain a decay rate by controlling the discrete relative entropy with the entropy production. This result is proved under assumptions of existence and uniform-in-time L^∞ estimates for numerical solutions, which are then discussed.

7.2. A posteriori analysis and computational optimization

In 2015, E. Creusé *et al.* have developed a posteriori error estimators for the harmonic potential formulations of the Maxwell system, in order to simulate eddy-current problems arising in the context of quasi-static approximations. The originality of our contribution is to provide estimators with sharp bounds and explicit constants. It was achieved by solving in the same time the so-called "$\mathbf{A}/\varphi$" and "$\mathbf{T}/\Omega$" potential formulations [38]. If this way to proceed was already known and usually used for stationary problems, the extension to harmonic ones constitutes the novelty of our contribution. It was in particular necessary to prove some superconvergence properties of additional terms. The reliability as well as the local efficiency of the derived estimator have been established without any generic constant, and numerical tests clearly illustrate their optimal behavior, from academic benchmarks to more industrial ones.

Another track to optimize the computational effort consists in refining and coarsening the model. This approach is based on the following ansatz : the more the model is complex, the more expensive are the computations. This approach was used by F. Filbet and T. Rey in [23] to simulate kinetic equations, the kinetic equations being replaced by cheaper hydrodynamic limits when it is relevant. The same idea was used in H. Mathis *et al.* [27] in order to simulate complex flows modeled by hyperbolic systems with relaxation. A rigorous error analysis of such a model adaptation procedure was performed on a simplified model by C. Cancès *et al.* in [13].

7.3. Modeling and numerical simulation of complex fluids

Recently, C. Calgaro *et al.* compared some very recent numerical schemes for the resolution of incompressible variable density flows; namely an Hybrid Finite Volume/Finite Element scheme, and a Discrete Duality Finite Volume one [34]. This work was performed in collaboration with the Inria team COFFEE (Inria Nice Sophia-Antipolis). In addition to this original and attentive comparison, our main contribution has been to improve the way to implement the Neumann boundary condition on the density, when a second-order accurate scheme is considered in space. Indeed, for some critical situations such as the simulation of Rayleigh-Taylor instabilities using unstructured meshes, this point is crucial to avoid unphysical numerical instabilities in the vicinity of the boundaries corresponding to symmetric axis. The obtained results are very promising, and constitute an important step towards the simulation of more complex models on which we are working at the moment.

In [25], M. Gisclon and I. Lacroix-Violet consider the barotropic compressible quantum Navier-Stokes equations with a linear density dependent viscosity and its limit when the scaled Planck constant vanish. Following recent works on degenerate compressible Navier-Stokes equations, we prove the global existence of weak solutions by the use of a singular pressure close to vacuum. With such singular pressure, we can use the standard definition of global weak solutions which also allows to justify the limit when the scaled Planck constant denoted by ε tends to 0.

The H-theorem, originally derived at the level of the Boltzmann nonlinear kinetic equation for a dilute gas undergoing elastic collisions, strongly constrains the velocity distribution of the gas to evolve irreversibly towards equilibrium. As such, the theorem could not be generalized to account for dissipative systems: the conservative nature of collisions is an essential ingredient in the standard derivation. The work [24] gives the first strong numerical evidences, along with a proof for a simplified model, of dissipation of the Boltzmann entropy (the so-called H-theorem) for solutions to the granular gases equation. This dissipative kinetic equation describes the non-equilibrium behavior of a gas composed of macroscopic particles, namely complex fluids such as avalanches, pollens flows or planetary rings.

7.4. Theoretical and numerical analysis of corrosion models

The Diffusion Poisson Coupled Model [1] is a model of iron based alloy in a nuclear waste repository. It describes the growth of an oxide layer in this framework. The system is made of a Poisson equation on the electrostatic potential and convection-diffusion equations on the densities f charge carriers (electrons, ferric cations and oxygen vacancies), supplemented with coupled Robin boundary conditions. The DPCM model also takes into account the growth of the oxide host lattice and its dissolution, leading to moving boundary equations.

In [19], C. Chainais-Hillairet and I. Lacroix-Violet consider a simplified version of this model, where only two charge carriers are taken into account and where there is no evolution of the layer thickness. They prove the existence of a solution for the time-dependent simplified model.

P.-L. Colin, C.Chainais-Hillairet and I. Lacroix-Violet have performed in [16] the numerical analysis of the numerical scheme presented in [2] for the same model. The scheme is a Euler implicit in time scheme with Scharfetter-Gummel approximation of the convection-diffusion fluxes. They prove existence of a solution to the scheme, a priori estimates satisfied by the solution and convergence of the numerical scheme to a weak solution of the corrosion model.

Numerical experiments done for the simulation of the full DPCM model with moving boundaries shows the convergence in time towards a pseudo-steady-state. C. Chainais-Hillairet and T. O. Gallouët show in [37] the existence of pseudo-stationary solutions for some simplified versions of the DPCM model. They also propose a new scheme in order to compute directly this pseudo-steady-state. Numerical experiments show the efficiency of this method.

7.5. **Variational modeling and analysis**

Bose-Einstein condensates are a unique way to observe quantum effects at a (relatively) large scale. The fundamental states of such condensates are obtained as minimizers of a Gross-Pitaievskii functional. In [39], M. Goldman and B. Merlet consider the case of a two component Bose-Einstein condensate in the strong segregation regime (the energy favors spatial segregation of the two different Boson species). They identify two different regimes in the strong segregation and small healing length limit. In one of these regimes, the relevant limit is an interesting weighted isoperimetric problem which explains some of the numerical simulations of [63].

In [14], C. Cancès *et al.* show that the equations that are classically used for modeling the motion of two incompressible immiscible phases in a porous medium can be formally reinterpreted as the gradient flow of the free energy in a degenerated geometry closely related to the Wasserstein metric. This extends to realistic models the seminal approach [65] and the more recent one [64].

7. New Results

7.1. Tools for understanding evolution

Automatic Detection of System-Specific Conventions. In Apache Ant, a convention to improve maintenance was introduced in 2004 stating a new way to close files instead of the Java generic InputStream.close(). Yet, six years after its introduction, this convention was still not generally known to the developers. Two existing solutions could help in these cases. First, one can deprecate entities, but, in our example, one can hardly deprecate Java's method. Second, one can create a system-specific rule to be automatically enforced. In a preceding publication , we showed that system-specific rules are more likely to be noticed by developers than generic ones. However, in practice, developers rarely create specific rules. We therefore propose to free the developers from the need to create rules by automatically detecting such conventions from source code repositories. This is done by mining the change history of the system to discover similar changes being applied over several revisions. The proposed approach is applied to real-world systems, and the extracted rules are validated with the help of experts. The results show that many rules are in fact relevant for the experts. [16]

DeltaImpactFinder. In software development, version control systems (VCS) provide branching and merging support tools. Such tools are popular among developers to concurrently change a code-base in separate lines and reconcile their changes automatically afterwards. However, two changes that are correct independently can introduce bugs when merged together. We call semantic merge conflicts this kind of bugs. Change impact analysis (CIA) aims at estimating the effects of a change in a codebase. We propose to detect semantic merge conflicts using CIA. On a merge, DELTAIMPACTFINDER analyzes and compares the impact of a change in its origin and destination branches. We call the difference between these two impacts the delta-impact. If the delta-impact is empty, then there is no indicator of a semantic merge conflict and the merge can continue automatically. Otherwise, the delta-impact contains what are the sources of possible conflicts. [26]

OrionPlanning. Many techniques have been proposed in the literature to support architecture definition, conformance, and analysis. However, there is a lack of adoption of such techniques by the industry. Previous work have analyzed this poor support. Specifically, former approaches lack proper analysis techniques (e.g., detection of architectural inconsistencies), and they do not provide extension and addition of new features. We present ORIONPLANNING, a prototype tool to assist refactorings at large scale. The tool provides support for model-based refactoring operations. These operations are performed in an interactive visualization. The contributions of the tool consist in: (i) providing iterative modifications in the architecture, and (ii) providing an environment for architecture inspection and definition of dependency rules. [37]

Recording and Replaying System-Specific Conventions. During its lifetime, a software system is under continuous maintenance to remain useful. Maintenance can be achieved in activities such as adding new features, fixing bugs, improving the system's structure, or adapting to new APIs. In such cases, developers sometimes perform sequences of code changes in a systematic way. These sequences consist of small code changes (e.g., create a class, then extract a method to this class), which are applied to groups of related code entities (e.g., some of the methods of a class). MacroRecorder is a proof-of-concept tool that records a sequence of code changes, then it allows the developer to generalize this sequence in order to apply it in other code locations. The evaluation is based on previous work on repetitive code changes related to rearchitecting. MacroRecorder was able to replay 92% of the examples, which consisted in up to seven code entities modified up to 66 times. The generation of a customizable, large-scale transformation operator has the potential to efficiently assist code maintenance. [39], [38]

7.2. Software Quality: Taming Software Evolution

Software metrics do not predict the health of a project. More and more companies would like to mine software data with the goal of assessing the health of their software projects. The hope is that some software

metrics could be tracked to predict failure risks or confirm good health. If a factor of success was found, projects failures could be anticipated and early actions could be taken by the organisation to help or to monitor closely the project, allowing one to act in a preventive mode rather than a curative one. We were called by a major IT company to fulfil this goal. We conducted a study to check whether software metrics can be related to project failure. The study was both theoretic with a review of literature on the subject, and practical with mining past projects data and interviews with project managers. We found that metrics used in practice are not reliable to assess project outcome. [22]

How Do Developers React to API Evolution? Software engineering research now considers that no system is an island, but it is part of an ecosystem involving other systems, developers, users, hardware,. .. When one system (e.g., a framework) evolves, its clients often need to adapt. Client developers might need to adapt to functionalities, client systems might need to be adapted to a new API, client users might need to adapt to a new User Interface. The consequences of such changes are yet unclear, what proportion of the ecosystem might be expected to react, how long might it take for a change to diffuse in the ecosystem, do all clients react in the same way? We report on an exploratory study aimed at observing API evolution and its impact on a large-scale software ecosystem, Pharo, which has about 3,600 distinct systems, more than 2,800 contributors, and six years of evolution. We analyze 118 API changes and answer research questions regarding the magnitude, duration, extension, and consistency of such changes in the ecosystem. The results of this study help to characterize the impact of API evolution in large software ecosystems, and provide the basis to better understand how such impact can be alleviated. [27]

Does JavaScript software embrace classes? JavaScript is the de facto programming language for the Web. It is used to implement mail clients, office applications, or IDEs, that can weight hundreds of thousands of lines of code. The language itself is prototype based, but to master the complexity of their application, practitioners commonly rely on some informal class abstractions. This practice has never been the target of empirical investigations in JavaScript. Yet, understanding it would be key to adequately tune programming environments and structure libraries such as they are accessible to programmers. We report a large and in-depth study to understand how class emulation is employed in JavaScript applications. We propose a strategy to statically detect class-based abstractions in the source code of JavaScript systems. We used this strategy in a dataset of 50 popular JavaScript applications available from GitHub. We found systems structured around hundreds of classes, suggesting that JavaScript developers are standing on traditional class-based abstractions to tackle the growing complexity of their systems. [28]

7.3. Software Quality: History and Changes

Mining Architectural Violations from Version History. Software architecture conformance is a key software quality control activity that aims to reveal the progressive gap normally observed between concrete and planned software architectures. However, formally specifying an architecture can be difficult, as it must be done by an expert of the system having a high level understanding of it. We present a lightweighted approach for architecture conformance based on a combination of static and historical source code analysis. The proposed approach relies on four heuristics for detecting absences (something expected was not found) and divergences (something prohibited was found) in source code based architectures. We also present an architecture conformance process based on the proposed approach. We followed this process to evaluate the architecture of two industrial-strength information systems, achieving an overall precision of 62.7% and 53.8%. We also evaluated our approach in an open-source information retrieval library, achieving an overall precision of 59.2%. We envision that an heuristic-based approach for architecture conformance can be used to rapidly raise architectural warnings, without deeply involving experts in the process. [17]

Untangling Fine-Grained Code Changes. After working for some time, developers commit their code changes to a version control system. When doing so, they often bundle unrelated changes (e.g., bug fix and refactoring) in a single commit, thus creating a so-called tangled commit. Sharing tangled commits is problematic because it makes review, reversion, and integration of these commits harder and historical analyses of the project less reliable. Researchers have worked at untangling existing commits, i.e., finding which part of a commit relates to which task. We contribute to this line of work in two ways: (1) A publicly available

dataset of untangled code changes, created with the help of two developers who accurately split their code changes into self contained tasks over a period of four months; (2) a novel approach, EpiceaUntangler, to help developers share untangled commits (aka. atomic commits) by using fine-grained code change information. EpiceaUntangler is based and tested on the publicly available dataset, and further evaluated by deploying it to 7 developers, who used it for 2 weeks. We recorded a median success rate of 91% and average one of 75%, in automatically creating clusters of untangled fine-grained code changes. [25]

Developers' Perception of Co-Change Patterns: An Empirical Study. Co-change clusters are groups of classes that frequently change together. They are proposed as an alternative modular view, which can be used to assess the traditional decomposition of systems in packages. To investigate developer's perception of co-change clusters, we report a study with experts on six systems, implemented in two languages. We mine 102 co-change clusters from the version history of such systems, which are classified in three patterns regarding their projection to the package structure: Encapsulated, Crosscutting, and Octopus. We then collect the perception of expert developers on such clusters, aiming to ask two central questions: (a) what concerns and changes are captured by the extracted clusters? (b) do the extracted clusters reveal design anomalies? We conclude that Encapsulated Clusters are often viewed as healthy designs and that Crosscutting Clusters tend to be associated to design anomalies. Octopus Clusters are normally associated to expected class distributions, which are not easy to implement in an encapsulated way, according to the interviewed developers. [40]

7.4. Dynamic Languages: Debugging

Practical domain-specific debuggers. Understanding the run-time behavior of software systems can be a challenging activity. Debuggers are an essential category of tools used for this purpose as they give developers direct access to the running systems. Nevertheless, traditional debuggers rely on generic mechanisms to introspect and interact with the running systems, while developers reason about and formulate domain-specific questions using concepts and abstractions from their application domains. This mismatch creates an abstraction gap between the debugging needs and the debugging support leading to an inefficient and error-prone debugging effort, as developers need to recover concrete domain concepts using generic mechanisms. To reduce this gap, and increase the efficiency of the debugging process, we propose a framework for developing domain-specific debuggers, called the Moldable Debugger, that enables debugging at the level of the application domain. The Moldable Debugger is adapted to a domain by creating and combining domain-specific debugging operations with domain-specific debugging views, and adapts itself to a domain by selecting, at run time, appropriate debugging operations and views. To ensure the proposed model has practical applicability (i.e., can be used in practice to build real debuggers), we discuss, from both a performance and usability point of view, three implementation strategies. We further motivate the need for domain-specific debugging, identify a set of key requirements and show how our approach improves debugging by adapting the debugger to several domains. [14]

Mercury: Properties and Design of a Remote Debugging Solution using Reflection. Remote debugging facilities are a technical necessity for devices that lack appropriate input/output interfaces (display, keyboard, mouse) for programming (e.g., smartphones, mobile robots) or are simply unreachable for local development (e.g., cloud-servers). Yet remote debugging solutions can prove awkward to use due to re-deployments. Empirical studies show us that on average 10.5 minutes per coding hour (over five 40-hour work weeks per year) are spent for redeploying applications (including re-deployments during debugging). Moreover current solutions lack facilities that would otherwise be available in a local setting because it is difficult to reproduce them remotely. Our work identifies three desirable properties that a remote debugging solution should exhibit, namely: run-time evolution, semantic instrumentation and adaptable distribution. Given these properties we propose and validate Mercury, a remote debugging model based on reflection. Mercury supports run-time evolution through a causally connected remote meta-level, semantic instrumentation through the reification of the underlying execution environment and adaptable distribution through a modular architecture of the debugging middleware. [19]

7.5. Reconciling Dynamic Languages and Isolation

Handles. Controlling object graphs and giving specific semantics to references (such as read-only, ownership, scoped sharing) have been the focus of a large body of research in the context of static type systems. Controlling references to single objects and to graphs of objects is essential to build more secure systems, but is notoriously hard to achieve in the absence of static type systems. In this article we embrace this challenge by proposing a solution to the following question: What is an underlying mechanism that can support the definition of properties (such as revocable, read-only, lent) at the reference level in the absence of a static type system? We present handles: first-class references that propagate behavioral change dynamically to the object subgraph during program execution. In this article we describe handles and show how handles support the implementation of read-only references and revocable references. Handles have been fully implemented by modifying an existing virtual machine and we report their costs. [13]

Delegation Proxies. Scoping behavioral variations to dynamic extents is useful to support non-functional concerns that otherwise result in cross-cutting code. Unfortunately, such forms of scoping are difficult to obtain with traditional reflection or aspects. We propose delegation proxies, a dynamic proxy model that supports behavioral intercession through the interception of various interpretation operations. Delegation proxies permit different behavioral variations to be easily composed together. We show how delegation proxies enable behavioral variations that can propagate to dynamic extents. We demonstrate our approach with examples of behavioral variations scoped to dynamic extents that help simplify code related to safety, reliability, and monitoring. [21]

Access Control to Reflection with Object Ownership. Reflection is a powerful programming language feature that enables language extensions, generic code, dynamic analyses , development tools, etc. However, uncontrolled reflection breaks object encapsulation and considerably increases the attack surface of programs e.g., malicious libraries can use reflection to attack their client applications. To bring reflection and object encapsulation back together, we use dynamic object ownership to design an access control policy to reflective operations. This policy grants objects full reflective power over the objects they own but limited reflective power over other objects. Code is still able to use advanced reflective operations but reflection cannot be used as an attack vector anymore. [41]

7.6. Tailoring Applications and boostrapping

Virtualization Support for Dynamic Core Library Update. Dynamically updating language runtime and core libraries such as collections and threading is challenging since the update mechanism uses such libraries at the same time that it modifies them. To tackle this challenge, we present Dynamic Core Library Update (DCU) as an extension of Dynamic Software Update (DSU) and our approach based on a virtualization architecture. Our solution supports the update of core libraries as any other normal library, avoiding the circular dependencies between the updater and the core libraries. Our benchmarks show that there is no evident performance overhead in comparison with a default execution. Finally, we show that our approach can be applied to real life scenario by introducing a critical update inside a web application with 20 simulated concurrent users. [34]

Bootstrapping Infrastructure. Bootstrapping is well known in the context of compilers, where a bootstrapped compiler can compile its own source code. Bootstrapping is a beneficial engineering practice because it raises the level of abstraction of a program making it easier to understand, optimize, evolve, etc. Bootstrapping a reflective object-oriented language is however more challenging , as we need also to initialize the runtime of the language with its initial objects and classes besides writing its compiler. We present a novel bootstrapping infrastructure for Pharo-like languages that allows us to easily extend and modify such languages. Our bootstrapping process relies on a first-class runtime. A first-class runtime is a meta-object that represents a program's runtime and provides a MOP to easily load code into it and manipulate its objects. It decou-ples the virtual machine (VM) and language concerns by introducing a clear VM-language interface. Using this process, we show how we succeeded to bootstrap a Smalltalk-based language named Candle and then extend it with traits in less than 250 lines of high-level Smalltalk code. We also show how we can bootstrap with minimal effort two other languages (Pharo and MetaTalk) with similar execution semantics but different object models. [35]

7.7. Dynamic Languages: Virtual Machines

Towards Fully Reflective Environments. Modern development environments promote live programming (LP) mechanisms because it enhances the development experience by providing instantaneous feedback and interaction with live objects. LP is typically supported with advanced reflective techniques within dynamic languages. These languages run on top of Virtual Machines (VMs) that are built in a static manner so that most of their components are bound at compile time. As a consequence, VM developers are forced to work using the traditional edit-compile-run cycle, even when they are designing LP-supporting environments. We explore the idea of bringing LP techniques to VM development to improve the observabil-ity, evolution and adaptability of VMs at run-time. We define the notion of fully reflective execution environments , systems that provide reflection not only at the application level but also at the level of the execution environment (EE). We characterize such systems, propose a design, and present Mate v1, a prototypical implementation. Based on our prototype, we analyze the feasibility and applicability of incorporating reflective capabilities into different parts of EEs. Furthermore, the evaluation demonstrates the opportunities such reflective capabilities provide for unanticipated dynamic adaptation scenarios, benefiting thus, a wider range of users. [23]

Tracing vs. Partial Evaluation. Tracing and partial evaluation have been proposed as meta-compilation techniques for interpreters to make just-in-time compilation language-independent. They promise that programs executing on simple interpreters can reach performance of the same order of magnitude as if they would be executed on state-of-the-art virtual machines with highly optimizing just-in-time compilers built for a specific language. Tracing and partial evaluation approach this meta-compilation from two ends of a spectrum, resulting in different sets of tradeoffs. This study investigates both approaches in the context of self-optimizing interpreters, a technique for building fast abstract-syntax-tree interpreters. Based on RPython for tracing and Truffle for partial evaluation, we assess the two approaches by comparing the impact of various optimizations on the performance of an interpreter for SOM, an object-oriented dynamically-typed language. The goal is to determine whether either approach yields clear performance or engineering benefits. We find that tracing and partial evaluation both reach roughly the same level of performance. SOM based on meta-tracing is on average 3x slower than Java, while SOM based on partial evaluation is on average 2.3x slower than Java. With respect to the engineering, tracing has however significant benefits, because it requires language implementers to apply fewer optimizations to reach the same level of performance. [29]

Zero-Overhead Metaprogramming. Runtime metaprogramming enables many useful applications and is often a convenient solution to solve problems in a generic way, which makes it widely used in frameworks, middleware, and domain-specific languages. However, powerful metaobject protocols are rarely supported and even common concepts such as reflective method invocation or dynamic proxies are not optimized. Solutions proposed in literature either restrict the metaprogramming capabilities or require application or library developers to apply performance improving techniques. For overhead-free runtime metaprogramming, we demonstrate that dispatch chains, a generalized form of polymorphic inline caches common to self-optimizing interpreters, are a simple optimization at the language-implementation level. Our evaluation with self-optimizing interpreters shows that unrestricted metaobject protocols can be realized for the first time without runtime overhead, and that this optimization is applicable for just-in-time compilation of interpreters based on meta-tracing as well as partial evaluation. In this context, we also demonstrate that optimizing common reflective operations can lead to significant performance improvements for existing applications [30].

A Partial Read Barrier for Efficient Support of Live Object-oriented Programming. Live programming, originally introduced by Smalltalk and Lisp, and now gaining popularity in contemporary systems such as Swift, requires on-the-fly support for object schema migration, such that the layout of objects may be changed while the program is at one and the same time being run and developed. In Smalltalk schema migration is supported by two primitives, one that answers a collection of all instances of a class, and one that exchanges the identities of pairs of objects, called the become primitive. Existing instances are collected, copies using the new schema created, state copied from old to new, and the two exchanged with become, effecting the schema migration. Historically the implementation of become has either required an extra level of indirection between an object's address and its body, slowing down slot access, or has required a sweep of all objects, a very slow operation on large heaps. Spur, a new object representation and memory manager for Smalltalk-like

languages, has neither of these deficiencies. It uses direct pointers but still provides a fast become operation in large heaps, thanks to forwarding objects that when read conceptually answer another object and a partial read barrier that avoids the cost of explicitly checking for forwarding objects on the vast majority of object accesses [31].

<center>**SEQUEL Project-Team**</center>

7. New Results

7.1. Decision-making Under Uncertainty

7.1.1. Reinforcement Learning

Nonparametric multiple change point estimation in highly dependent time series [7]

Given a heterogeneous time-series sample, the objective is to find points in time, called change points, where the probability distribution generating the data has changed. The data are assumed to have been generated by arbitrary unknown stationary ergodic distributions. No modelling, independence or mixing assumptions are made. A novel, computationally efficient, nonparametric method is proposed, and is shown to be asymptotically consistent in this general framework. The theoretical results are complemented with experimental evaluations.

Explore no more: Improved high-probability regret bounds for non-stochastic bandits [26]

This work addresses the problem of regret minimization in non-stochastic multi-armed bandit problems, focusing on performance guarantees that hold with high probability. Such results are rather scarce in the literature since proving them requires a large deal of technical effort and significant modifications to the standard, more intuitive algorithms that come only with guarantees that hold on expectation. One of these modifications is forcing the learner to sample arms from the uniform distribution at least $\Omega(\sqrt{T})$ times over T rounds, which can adversely affect performance if many of the arms are suboptimal. While it is widely conjectured that this property is essential for proving high-probability regret bounds, we show in this paper that it is possible to achieve such strong results without this undesirable exploration component. Our result relies on a simple and intuitive loss-estimation strategy called Implicit eXploration (IX) that allows a remarkably clean analysis. To demonstrate the flexibility of our technique, we derive several improved high-probability bounds for various extensions of the standard multi-armed bandit framework. Finally, we conduct a simple experiment that illustrates the robustness of our implicit exploration technique.

First-order regret bounds for combinatorial semi-bandits [27]

We consider the problem of online combinatorial optimization under semi-bandit feedback, where a learner has to repeatedly pick actions from a combinatorial decision set in order to minimize the total losses associated with its decisions. After making each decision, the learner observes the losses associated with its action, but not other losses. For this problem, there are several learning algorithms that guarantee that the learner's expected regret grows as $O(\sqrt{T})$ with the number of rounds T. In this paper, we propose an algorithm that improves this scaling to $O(\sqrt{L * T})$, where L * T is the total loss of the best action. Our algorithm is among the first to achieve such guarantees in a partial-feedback scheme, and the first one to do so in a combinatorial setting.

Random-Walk Perturbations for Online Combinatorial Optimization [4]

We study online combinatorial optimization problems where a learner is interested in minimizing its cumulative regret in the presence of switching costs. To solve such problems, we propose a version of the follow-the-perturbed-leader algorithm in which the cumulative losses are perturbed by independent symmetric random walks. In the general setting, our forecaster is shown to enjoy near-optimal guarantees on both quantities of interest, making it the best known efficient algorithm for the studied problem. In the special case of prediction with expert advice, we show that the forecaster achieves an expected regret of the optimal order $O(\sqrt{n \log N})$ where n is the time horizon and N is the number of experts, while guaranteeing that the predictions are switched at most $O(\sqrt{n \log N})$ times, in expectation.

Qualitative Multi-Armed Bandits: A Quantile-Based Approach [32]

We formalize and study the multi-armed bandit (MAB) problem in a generalized stochastic setting, in which rewards are not assumed to be numerical. Instead, rewards are measured on a qualitative scale that allows for comparison but invalidates arithmetic operations such as averaging. Correspondingly, instead of characterizing an arm in terms of the mean of the underlying distribution, we opt for using a quantile of that distribution as a representative value. We address the problem of quantile-based online learning both for the case of a finite (pure exploration) and infinite time horizon (cumulative regret minimization). For both cases, we propose suitable algorithms and analyze their properties. These properties are also illustrated by means of first experimental studies.

Predicting the outcomes of every process for which an asymptotically accurate stationary predictor exists is impossible [30]

The problem of prediction consists in forecasting the conditional distribution of the next outcome given the past. Assume that the source generating the data is such that there is a stationary predictor whose error converges to zero (in a certainsense). The question is whether there is a universal predictor for all such sources, that is, a predictor whose error goes to zero if any of the sources that have this property is chosen to generate the data. This question is answered in the negative, contrasting a number of previously established positive results concerning related but smaller sets of processes.

Improved Regret Bounds for Undiscounted Continuous Reinforcement Learning [22]

We consider the problem of undiscounted reinforcement learning in continuous state space. Regret bounds in this setting usually hold under various assumptions on the structure of the reward andtransition function. Under the assumption that the rewards andtransition probabilities are Lipschitz, for 1-dimensional state space a regret bound of $\widetilde{O}(T^{\frac{3}{4}})$ after any T steps has been given by.Here we improve upon this result by using non-parametric kernel density estimation for estimating the transition probability distributions,and obtain regret bounds that depend on the smoothness of the transition probability distributions.In particular, under the assumption that the transition probability functions are smoothly differentiable, the regret bound is shown to be $\widetilde{O}(T^{\frac{2}{3}})$ asymptotically for reinforcement learning in 1-dimensional state space. Finally, we also derive improved regret bounds for higher dimensional state space.

Maximum Entropy Semi-Supervised Inverse Reinforcement Learning [9]

A popular approach to apprenticeship learning (AL) is to formulate itas an inverse reinforcement learning (IRL) problem. The MaxEnt-IRL algorithm successfully integrates the maximum entropy principleinto IRL and unlike its predecessors, it resolves theambiguity arising from the fact that a possibly large number of policies couldmatch the expert's behavior. In this paper, we study an AL setting in which inaddition to the expert's trajectories,a number of unsupervised trajectories is available. We introduce MESSI,a novel algorithm that combines MaxEnt-IRLwith principles coming from semi-supervised learning. In particular, MESSIintegrates the unsupervised data intothe MaxEnt-IRL framework using a pairwise penalty on trajectories. Empirical-results in a highway driving and grid-world problems indicate that MESSI is able to take advantage of the unsupervised trajectories and improve the performance ofMaxEnt-IRL.

Direct Policy Iteration with Demonstrations [12]

We consider the problem of learning the optimal policy of an unknown Markov decision process (MDP) when expert demonstrations are available along with interaction samples. We build on classification-based policy iteration to perform a seamless integration of interaction and expert data, thus obtaining an algorithm which can benefit from both sources of information at the same time. Furthermore , we provide a full theoretical analysis of the performance across iterations providing insights on how the algorithm works. Finally, we report an empirical evaluation of the algorithm and a comparison with the state-of-the-art algorithms.

Approximate Modified Policy Iteration and its Application to the Game of Tetris [8]

Modified policy iteration (MPI) is a dynamic programming (DP) algorithm that contains the two celebrated policy and value iteration methods. Despite its generality, MPI has not been thoroughly studied, especially its approximation form which is used when the state and/or action spaces are large or infinite. In this

paper, we propose three implementations of approximate MPI (AMPI) that are extensions of the well-known approximate DP algorithms: fitted-value iteration, fitted-Q iteration, and classification-based policy iteration. We provide error propagation analysis that unify those for approximate policy and value iteration. We develop the finite-sample analysis of these algorithms, which highlights the influence of their parameters. In the classification-based version of the algorithm (CBMPI), the analysis shows that MPI's main parameter controls the balance between the estimation error of the classifier and the overall value function approximation. We illustrate and evaluate the behavior of these new algorithms in the Mountain Car and Tetris problems. Remarkably, in Tetris, CBMPI outperforms the existing DP approaches by a large margin, and competes with the current state-of-the-art methods while using fewer samples.

7.1.2. *Multi-arm Bandit Theory*

Simple regret for infinitely many armed bandits [11]

We consider a stochastic bandit problem with infinitely many arms. In this setting, the learner has no chance of trying all the arms even once and has to dedicate its limited number of samples only to a certain number of arms. All previous algorithms for this setting were designed for minimizing the cumulative regret of the learner. In this paper, we propose an algorithm aiming at minimizing the simple regret. As in the cumulative regret setting of infinitely many armed bandits , the rate of the simple regret will depend on a parameter β characterizing the distribution of the near-optimal arms. We prove that depending on β, our algorithm is minimax optimal either up to a multiplicative constant or up to a log(n) factor. We also provide extensions to several important cases: when β is unknown, in a natural setting where the near-optimal arms have a small variance , and in the case of unknown time horizon.

Black-box optimization of noisy functions with unknown smoothness [20]

We study the problem of black-box optimization of a function f of any dimension, given function evaluations perturbed by noise. The function is assumed to be locally smooth around one of its global optima, but this smoothness is unknown. Our contribution is an adaptive optimization algorithm, POO or parallel optimistic optimization, that is able to deal with this setting. POO performs almost as well as the best known algorithms requiring the knowledge of the smoothness. Furthermore, POO works for a larger class of functions than what was previously considered, especially for functions that are difficult to optimize, in a very precise sense. We provide a finite-time analysis of POO's performance, which shows that its error after n evaluations is at most a factor of $\sqrt{\ln n}$ away from the error of the best known optimization algorithms using the knowledge of the smoothness.

Cheap Bandits [21]

We consider stochastic sequential learning problems where the learner can observe the average reward of several actions. Such a setting is interesting in many applications involving monitoring and surveillance, where the set of the actions to observe represent some (geographical) area. The importance of this setting is that in these applications , it is actually cheaper to observe average reward of a group of actions rather than the reward of a single action. We show that when the reward is smooth over a given graph representing the neighboring actions, we can maximize the cumulative reward of learning while minimizing the sensing cost. In this paper we propose CheapUCB, an algorithm that matches the regret guarantees of the known algorithms for this setting and at the same time guarantees a linear cost again over them. As a by-product of our analysis , we establish a (p dT) lower bound on the cumulative regret of spectral bandits for a class of graphs with effective dimension d.

Truthful Learning Mechanisms for Multi–Slot Sponsored Search Auctions with Externalities [5]

Sponsored Search Auctions (SSAs) constitute one of the most successful applications of microeconomic mechanisms. In mechanism design, auctions are usually designed to incentivize advertisers to bid their truthful valuations and, at the same time, to guarantee both the advertisers and the auctioneer a non–negative utility. Nonetheless, in sponsored search auctions, the Click–Through–Rates (CTRs) of the advertisers are often unknown to the auctioneer and thus standard truthful mechanisms cannot be directly applied and must be

paired with an effective learning algorithm for the estimation of the CTRs. This introduces the critical problem of designing a learning mechanism able to estimate the CTRs at the same time as implementing a truthful mechanism with a revenue loss as small as possible compared to the mechanism that can exploit the true CTRs. Previous work showed that, when dominant–strategy truthfulness is adopted, in single–slot auctions the problem can be solved using suitable exploration–exploitation mechanisms able to achieve a cumulative regret (on the auctioneer's revenue) of order $O(T^{(2/3)})$, where T is the number of times the auction is repeated. It is also known that, when truthfulness in expectation is adopted, a cumulative regret (over the social welfare) of order $O(T^{(1/2)})$ can be obtained. In this paper, we extend the results available in the literature to the more realistic case of multi–slot auctions. In this case, a model of the user is needed to characterize how the CTR of an ad changes as its position in the allocation changes. In particular, we adopt the cascade model, one of the most popular models for sponsored search auctions, and we prove a number of novel upper bounds and lower bounds on both auctioneer's revenue loss and social welfare w.r.t. to the Vickrey–Clarke–Groves (VCG) auction. Furthermore, we report numerical simulations investigating the accuracy of the bounds in predicting the dependency of the regret on the auction parameters.

A Relative Exponential Weighing Algorithm for Adversarial Utility-based Dueling Bandits   *[37]*

We study the K-armed dueling bandit problem which is a variation of the classical Multi-Armed Bandit (MAB) problem in which the learner receives only relative feedback about the selected pairs of arms. We propose a new algorithm called Relative Exponential-weight algorithm for Exploration and Exploitation (REX3) to handle the adversarial utility-based formulation of this problem. This algorithm is a non-trivial extension of the Exponential-weight algorithm for Exploration and Exploitation (EXP3) algorithm. We prove a finite time expected regret upper bound of order O(sqrt(K ln(K)T)) for this algorithm and a general lower bound of order omega(sqrt(KT)). At the end, we provide experimental results using real data from information retrieval applications.

Simultaneous Optimistic Optimization on the Noiseless BBOB Testbed *[15]*

We experiment the SOO (Simultaneous Optimistic Optimization) global optimizer on the BBOB testbed. We report results for both the unconstrained-budget setting and the expensive setting, as well as a comparison with the DiRect algorithm to which SOO is mostly related. Overall, SOO is shown to perform rather poorly in the highest dimensions while agreeably exhibiting interesting performance for the most difficult functions, which is to be attributed to its global nature and to the fact that its design was guided by the goal of obtaining theoretically provable performance. The greedy exploration-exploitation sampling strategy underlying SOO design is also shown to be a viable alternative for the expensive setting which gives rooms for further improvements in this direction.

7.1.3. Recommendation systems

Bandits and Recommender Systems *[23]*

This paper addresses the on-line recommendation problem facing new users and new items; we assume that no information is available neither about users, nor about the items. The only source of information is a set of ratings given by users to some items. By on-line, we mean that the set of users, and the set of items, and the set of ratings is evolving along time and that at any moment, the recommendation system has to select items to recommend based on the currently available information, that is basically the sequence of past events. We also mean that each user comes with her preferences which may evolve along short and longer scales of time; so we have to continuously update their preferences. When the set of ratings is the only available source of information , the traditional approach is matrix factorization. In a decision making under uncertainty setting, actions should be selected to balance exploration with exploitation; this is best modeled as a bandit problem. Matrix factors provide a latent representation of users and items. These representations may then be used as contextual information by the bandit algorithm to select items. This last point is exactly the originality of this paper: the combination of matrix factorization and bandit algorithms to solve the on-line recommendation problem. Our work is driven by considering the recommendation problem as a feedback controlled loop. This leads to interactions between the representation learning, and the recommendation policy.

Collaborative Filtering as a Multi-Armed Bandit [35]

Recommender Systems (RS) aim at suggesting to users one or several items in which they might have interest. Following the feedback they receive from the user, these systems have to adapt their model in order to improve future recommendations. The repetition of these steps defines the RS as a sequential process. This sequential aspect raises an exploration-exploitation dilemma, which is surprisingly rarely taken into account for RS without contextual information. In this paper we present an explore-exploit collaborative filtering RS, based on Matrix Factor-ization and Bandits algorithms. Using experiments on artificial and real datasets, we show the importance and practicability of using sequential approaches to perform recommendation. We also study the impact of the model update on both the quality and the computation time of the recommendation procedure.

AUC Optimisation and Collaborative Filtering [39]

In recommendation systems, one is interested in the ranking of the predicted items as opposed to other losses such as the mean squared error. Although a variety of ways to evaluate rankings exist in the literature, here we focus on the Area Under the ROC Curve (AUC) as it widely used and has a strong theoretical underpinning. In practical recommendation, only items at the top of the ranked list are presented to the users. With this in mind, we propose a class of objective functions over matrix factorisations which primarily represent a smooth surrogate for the real AUC, and in a special case we show how to prioritise the top of the list. The objectives are differentiable and optimised through a carefully designed stochastic gradient-descent-based algorithm which scales linearly with the size of the data. In the special case of square loss we show how to improve computational complexity by leveraging previously computed measures. To understand theoretically the underlying matrix factorisation approaches we study both the consistency of the loss functions with respect to AUC, and generalisation using Rademacher theory. The resulting generalisation analysis gives strong motivation for the optimisation under study. Finally, we provide computation results as to the efficacy of the proposed method using synthetic and real data.

Collaborative Filtering with Localised Ranking [16]

In recommendation systems, one is interested in the ranking of the predicted items as opposed to other losses such as the mean squared error. Although a variety of ways to evaluate rankings exist in the literature, here we focus on the Area Under the ROC Curve (AUC) as it widely used and has a strong theoretical underpinning. In practical recommendation, only items at the top of the ranked list are presented to the users. With this in mind we propose a class of objective functions which primarily represent a smooth surrogate for the real AUC, and in a special case we show how to prioritise the top of the list. This loss is differentiable and is optimised through a carefully designed stochastic gradient-descent-based algorithm which scales linearly with the size of the data. We mitigate sample bias present in the data by sampling observations according to a certain power-law based distribution. In addition, we provide computation results as to the efficacy of the proposed method using synthetic and real data.

Collaborative Filtering with Stacked Denoising AutoEncoders and Sparse Inputs [36]

Neural networks have not been widely studied in Collaborative Filtering. For instance, no paper using neural networks was published during the Net-flix Prize apart from Salakhutdinov et al's work on Restricted Boltz-mann Machine (RBM) [14]. While deep learning has tremendous success in image and speech recognition, sparse inputs received less attention and remains a challenging problem for neural networks. Nonetheless, sparse inputs are critical for collaborative filtering. In this paper, we introduce a neural network architecture which computes a non-linear matrix factorization from sparse rating inputs. We show experimentally on the movieLens and jester dataset that our method performs as well as the best collaborative filtering algorithms. We provide an implementation of the algorithm as a reusable plugin for Torch [4], a popular neural network framework.

7.1.4. Nonparametric statistics of time series

The Replacement Bootstrap for Dependent Data [31]

Applications that deal with time-series data often require evaluating complex statistics for which each time series is essentially one data point. When only a few time series are available, bootstrap methods are used to generate additional samples that can be used to evaluate empirically the statistic of interest. In this work a novel bootstrap method is proposed, which is shown to have some asymptotic consistency guarantees under the only assumption that the time series are stationary and ergodic. This contrasts previously available results that impose mixing or finite-memory assumptions on the data. Empirical evaluation on simulated and real data, using a practically relevant and complex extrema statistic is provided.

7.1.5. *Imitation and Inverse Reinforcement Learning*

Inverse Reinforcement Learning in Relational Domains [24]

In this work, we introduce the first approach to the Inverse Reinforcement Learning (IRL) problem in relational domains. IRL has been used to recover a more compact representation of the expert policy leading to better generalization performances among different contexts. On the other hand, rela-tional learning allows representing problems with a varying number of objects (potentially infinite), thus provides more generalizable representations of problems and skills. We show how these different formalisms allow one to create a new IRL algorithm for relational domains that can recover with great efficiency rewards from expert data that have strong generalization and transfer properties. We evaluate our algorithm in representative tasks and study the impact of diverse experimental conditions such as : the number of demonstrations, knowledge about the dynamics, transfer among varying dimensions of a problem, and changing dynamics.

Imitation Learning Applied to Embodied Conversational Agents [29]

Embodied Conversational Agents (ECAs) are emerging as a key component to allow human interact with machines. Applications are numerous and ECAs can reduce the aversion to interact with a machine by providing user-friendly interfaces. Yet, ECAs are still unable to produce social signals appropriately during their interaction with humans, which tends to make the interaction less instinctive. Especially, very little attention has been paid to the use of laughter in human-avatar interactions despite the crucial role played by laughter in human-human interaction. In this paper, methods for predicting when and how to laugh during an interaction for an ECA are proposed. Different Imitation Learning (also known as Apprenticeship Learning) algorithms are used in this purpose and a regularized classification algorithm is shown to produce good behavior on real data.

7.1.6. *Stochastic Games*

Optimism in Active Learning [3]

Active learning is the problem of interactively constructing the training set used in classification in order to reduce its size. It would ideally successively add the instance-label pair that decreases the classification error most. However, the effect of the addition of a pair is not known in advance. It can still be estimated with the pairs already in the training set. The online minimization of the classification error involves a tradeoff between exploration and exploitation. This is a common problem in machine learning for which multiarmed bandit, using the approach of Optimism int the Face of Uncertainty, has proven very efficient these last years. This paper introduces three algorithms for the active learning problem in classification using Optimism in the Face of Uncertainty. Experiments lead on built-in problems and real world datasets demonstrate that they compare positively to state-of-the-art methods.

Bayesian Credible Intervals for Online and Active Learning of Classification Trees [13]

Classification trees have been extensively studied for decades. In the online learning scenario, a whole class of algorithms for decision trees has been introduced, called incremental decision trees. In the case where subtrees may not be discarded, an incremental decision tree can be seen as a sequential decision process, consisting in deciding to extend the existing tree or not. This problem involves an trade-off between exploration and exploitation, which is addressed in recent work with the use of Hoeffding's bounds. This paper proposes to use Bayesian Credible Intervals instead, in order to get the most out of the knowledge of the output's

distribution's shape. It also studies the case of Active Learning in such a tree following the Optimism in the Face of Uncertainty paradigm. Two novel algorithms are introduced for the online and active learning problems. Evaluations on real-world datasets show that these algorithms compare positively to state-of-the-art.

Optimism in Active Learning with Gaussian Processes [14]

In the context of Active Learning for classification, the classification error depends on the joint distribution of samples and their labels which is initially unknown. The minimization of this error requires estimating this distribution. Online estimation of this distribution involves a trade-off between exploration and exploitation. This is a common problem in machine learning for which multi-armed bandit theory, building upon Optimism in the Face of Uncertainty, has been proven very efficient these last years. We introduce two novel algorithms that use Optimism in the Face of Uncertainty along with Gaussian Processes for the Active Learning problem. The evaluation lead on real world datasets shows that these new algorithms compare positively to state-of-the-art methods.

Approximate Dynamic Programming for Two-Player Zero-Sum Markov Games [28]

This paper provides an analysis of error propagation in Approximate Dynamic Programming applied to zero-sum two-player Stochastic Games. We provide a novel and unified error propagation analysis in L p-norm of three well-known algorithms adapted to Stochastic Games (namely Approximate Value Iteration, Approximate Policy Iteration and Approximate Generalized Policy Iteratio,n). We show that we can achieve a stationary policy which is $2\gamma + (1-\gamma)$ 2-optimal, where is the value function approximation error and is the approximate greedy operator error. In addition , we provide a practical algorithm (AGPI-Q) to solve infinite horizon γ-discounted two-player zero-sum Stochastic Games in a batch setting. It is an extension of the Fitted-Q algorithm (which solves Markov Decisions Processes from data) and can be non-parametric. Finally, we demonstrate experimentally the performance of AGPI-Q on a simultaneous two-player game, namely Alesia.

7.2. Statistical analysis of time series

7.2.1. *Automata Learning*

Non-negative Spectral Learning for Linear Sequential Systems [18]

Method of moments (MoM) has recently become an appealing alternative to standard iterative approaches like Expectation Maximization (EM) to learn latent variable models. In addition, MoM-based algorithms come with global convergence guarantees in the form of finite sample bounds. However, given enough computation time, by using restarts and heuristics to avoid local optima, iterative approaches often achieve better performance. We believe that this performance gap is in part due to the fact that MoM-based algorithms can output negative probabilities. By constraining the search space, we propose a non-negative spectral algorithm (NNSpectral) avoiding computing negative probabilities by design. NNSpectral is compared to other MoM-based algorithms and EM on synthetic problems of the PAutomaC challenge. Not only, NNSpectral outperforms other MoM-based algorithms, but also, achieves very competitive results in comparison to EM.

Learning of scanning strategies for electronic support using predictive state representations [17]

In Electronic Support, a receiver must monitor a wide frequency spectrum in which threatening emitters operate. A common approach is to use sensors with high sensitivity but a narrow band-width. To maintain surveillance over the whole spectrum, the sensor has to sweep between frequency bands but requires a scanning strategy. Search strategies are usually designed prior to the mission using an approximate knowledge of illumination patterns. This often results in open-loop policies that cannot take advantage of previous observations. As pointed out in past researches, these strategies lack of robustness to the prior. We propose a new closed loop search strategy that learns a stochastic model of each radar using predic-tive state representations. The learning algorithm benefits from the recent advances in spectral learning and rank minimization using nuclear norm penalization.

Spectral learning with proper probabilities for finite state automation [19]

Probabilistic Finite Automaton (PFA), Probabilistic Finite State Transducers (PFST) and Hidden Markov Models (HMM) are widely used in Automatic Speech Recognition (ASR), Text-to-Speech (TTS) systems and Part Of Speech (POS) tagging for language mod-eling. Traditionally, unsupervised learning of these latent variable models is done by Expectation-Maximization (EM)-like algorithms, as the Baum-Welch algorithm. In a recent alternative line of work, learning algorithms based on spectral properties of some low order moments matrices or tensors were proposed. In comparison to EM, they are orders of magnitude faster and come with theoretical convergence guarantees. However, returned models are not ensured to compute proper distributions. They often return negative values that do not sum to one, limiting their applicability and preventing them to serve as an initialization to EM-like algorithms. In this paper, we propose a new spectral algorithm able to learn a large range of models constrained to return proper distributions. We assess its performances on synthetic problems from the PAutomaC challenge and real datasets extracted from Wikipedia. Experiments show that it outperforms previous spectral approaches as well as the Baum-Welch algorithm with random restarts, in addition to serve as an efficient initialization step to EM-like algorithms.

7.3. Statistical Learning and Bayesian Analysis

7.3.1. Prediction of Sequences of Structured and Unstructured Data

Operator-valued Kernels for Learning from Functional Response Data [6]

In this paper we consider the problems of supervised classification and regression in the case where attributes and labels are functions: a data is represented by a set of functions, and the label is also a function. We focus on the use of reproducing kernel Hilbert space theory to learn from such functional data. Basic concepts and properties of kernel-based learning are extended to include the estimation of function-valued functions. In this setting, the representer theorem is restated, a set of rigorously defined infinite-dimensional operator-valued kernels that can be valuably applied when the data are functions is described, and a learning algorithm for nonlinear functional data analysis is introduced. The methodology is illustrated through speech and audio signal processing experiments.

7.4. Applications

7.4.1. Software development

An Experimental Protocol for Analyzing the Accuracy of Software Error Impact Analysis [25]

In software engineering, error impact analysis consists in predicting the software elements (e.g. modules, classes, methods) potentially impacted by a change. Impact analysis is required to optimize the testing effort. In this paper we present a new protocol to analyze the accuracy of impact analysis. This protocol uses mutation testing to simulate changes that introduce errors. To this end, we introduce a variant of call graphs we name the "use graph" of a software which may be computed efficiently. We apply this protocol to two open-source projects and correctly predict the impact of 30

A Learning Algorithm for Change Impact Prediction: Experimentation on 7 Java Applications [41]

Change impact analysis consists in predicting the impact of a code change in a software application. In this paper, we take a learning perspective on change impact analysis and consider the problem formulated as follows. The artifacts that are considered are methods of object-oriented software; the change under study is a change in the code of the method, the impact is the test methods that fail because of the change that has been performed. We propose an algorithm, called LCIP that learns from past impacts to predict future impacts. To evaluate our system, we consider 7 Java software applications totaling 214,000+ lines of code. We simulate 17574 changes and their actual impact through code mutations, as done in mutation testing. We find that LCIP can predict the impact with a precision of 69

7.4.2. Spoken Dialogue Systems

Human-Machine Dialogue as a Stochastic Game [10]

In this paper, an original framework to model human-machine spoken dialogues is proposed to deal with co-adaptation between users and Spoken Dialogue Systems in non-cooperative tasks. The conversation is modeled as a Stochastic Game: both the user and the system have their own preferences but have to come up with an agreement to solve a non-cooperative task. They are jointly trained so the Dialogue Manager learns the optimal strategy against the best possible user. Results obtained by simulation show that non-trivial strategies are learned and that this framework is suitable for dialogue modeling.

SPIRALS Project-Team

7. New Results

7.1. Traceability of Concerns in Large Software Systems

In 2015, we obtained new results in the domain of the analysis of large software systems. The purpose is to be able to deal the complexity of such systems by slicing them depending on different concerns. The slicing enables to gain a view and a better understanding on how the concern evolves over time and through the different refinement layers of the software system. For that, we present a systematic approach based on model driven engineering and basic models of software components, in order to better manage software complexity and traceability of functional and non-functional requirements. We provide in particular three major contributions. First, we provide an integrated set of meta-models for describing the concerns of software requirements, software components, and traceability between the concerns and software components. By providing an abstract model, we are independent of any implementation and thus allow existing approaches relying on that model to expand their support. With the second contribution, we propose a formal support of our model to allow formal verification. We focus on temporal property verification. For this, our design model is translated into timed automata for which we can apply a timed model checker. Instead of using temporal logic, which is difficult to handle by non-experts, we use patterns of temporal properties. For each pattern, we propose timed automata that can be applied directly into a timed model checking tool. These timed automata are seen as observers or watch dogs that check the system under observation. Finally, with the last contribution, we propose a software component-based development and verification approach, called SARA, and included in V-lifecycle widely used in the railway domain. These contributions have been validated with case studies from the domain of railway control systems especially for the new European train control system ERTMS/ETCS. These results contribute to our objective on self-optimizing software systems (see Section 3.3) and are part of the PhD thesis by Marc Sango [13].

7.2. Automatic Analysis and Repair of Exception Bugs for Java Programs

In 2015, we obtained new results in the field of automated software repair, that is a new and emerging domain of software engineering. The goal of automatic repair is to increase the quality of software systems by automatizing tasks related to fixing of defects and bugs. The new results that we bring are related with the management of runtime exceptions. These results contribute to our objective on self-healing software systems (see Section 3.2) and are part of the PhD thesis by Benoit Cornu [11], defended on 26 November 2015. To improve the available information about exceptions, we have presented a characterization of the exceptions (expected or not, anticipated or not), and of their corresponding resilience mechanisms [16]. We have provided definitions about what is a bug when facing exceptions and what are the already-in-place corresponding resilience mechanisms. We have formalized two formal resilience properties: source-independence and pure-resilience as well as an algorithm to verify them. Then, we have presented two dynamic analysis techniques based on code transformation for analyzing exceptions. Casper is an approach to make bug fixing easier by providing information about the origin of null pointer dereferences. NpeFix is a system to tolerate null pointer dereferences. Both systems are empirically validated on real-world null dereference bugs from large-scale open-source projects