RESEARCH CENTER
Lille - Nord Europe

FIELD

Activity Report 2017

Section New Results

Edition: 2018-02-19

DOLPHIN Team

7. New Results

7.1. Fitness Landscape Analysis for Algorithm Understanding, Selection, Design and Configuration

Participants: Bilel Derbel, Arnaud Liefooghe (external collaborators: Sebastien Verel, Univ. Littoral, France; Hernan Aguirre, Fabio Daolio, Hugo Monzón, Miyako Sagawa and Kiyoshi Tanaka, Shinshu University, Japan)

Fitness landscape analysis of Pareto local search on bi-objective permutation flowshop scheduling problems. In [20], we study the difficulty of solving different bi-objective formulations of the permutation flowshop scheduling problem by adopting a fitness landscape analysis perspective. Our main goal is to shed the light on how different problem features can impact the performance of Pareto local search algorithms. Specifically, we conduct an empirical analysis addressing the challenging question of quantifying the individual effect and the joint impact of different problem features on the success rate of the considered approaches. Our findings support that multi-objective fitness landscapes enable to devise sound general-purpose features for assessing the expected difficulty in solving permutation flowshop scheduling problems, hence pushing a step towards a better understanding of the challenges that multi-objective randomized search heuristics have to face.

Landscape-aware automatic algorithm configuration. The proper setting of algorithm parameters is a well-known issue that gave rise to recent research investigations from the (offline) automatic algorithm configuration perspective. Besides, the characteristics of the target optimization problem is also a key aspect to elicit the behavior of a dedicated algorithm, and as often considered from a landscape analysis perspective. In [21], we show that fitness landscape analysis can open a whole set of new research opportunities for increasing the effectiveness of existing automatic algorithm configuration methods. Specifically, we show that using landscape features in iterated racing both (i) at the training phase, to compute multiple elite configurations explicitly mapped with different feature values, and (ii) at the production phase, to decide which configuration to use on a feature basis, provides significantly better results compared against the standard landscape-oblivious approach. Our first experimental investigations on NK-landscapes, considered as a benchmark family having controllable features in terms of ruggedness and neutrality, and tackled using a memetic algorithm with tunable population size and variation operators, show that a landscape-aware approach is a viable alternative to handle the heterogeneity of (black-box) combinatorial optimization problems.

Learning variable importance to guide recombination on many-objective optimization. There are numerous many-objective real-world problems in various application domains for which it is difficult or time-consuming to derive Pareto optimal solutions. In an evolutionary algorithm, variation operators such as recombination and mutation are extremely important to obtain an effective solution search. In [25], we study a machine learning-enhanced recombination that incorporates an intelligent variable selection method. The method is based on the importance of variables with respect to convergence to the Pareto front. We verify the performance of the enhanced recombination on benchmark test problems with three or more objectives using the many-objective evolutionary algorithm AϵSϵH as a baseline algorithm. Results show that variable importance can enhance the performance of many-objective evolutionary algorithms.

Closed state model for understanding the dynamics of multi-objective evolutionary algorithms. In [22], we propose the use of simple closed state models to capture, analyze and compare the dynamics of multi- and many-objective evolutionary algorithms. Two- and three-state models representing the composition of the instantaneous population are described and learned for representatives of the major approaches to multi-objective optimization, i.e. dominance, extensions of dominance, decomposition, and indicator algorithms. The model parameters are trained from data obtained running the algorithms with various population sizes on enumerable MNK-landscapes with 3, 4, 5 and 6 objectives. We show ways to interpret and use the model

parameter values in order to analyze the population dynamics according to selected features. For example, we are interested in knowing how parameter values change for a given population size with the increase of the number of objectives. We also show a graphical representation capturing in one graph how the parameters magnitude and sign relate to the connections between states.

7.2. Multi-objective Demand side management in smart grids

Participants: El-Ghazali Talbi (externel collaborators: Rachid Ellaia, Zineb Garroussi - Univ. Rabat Morocco)

Residential demand side management (DSM) is one of the most challenging topics in smart grids. In this work, a multi-objective model for the residential DSM is proposed. The smart home is composed of appliances, a battery and a photovoltaic panel. The resolution of this model is a matheuristic based on combining a multi-objective evolutionary algorithm and an exact liner programming solver (CPLEX). Candidate solutions in this hybrid approach are incompletely represented in the representation, and the exact solver is used as a decoder to determine the missing parts in an optimal way. In our case, hybridization involves solving a MILP sub-problem by CPLEX to manage the battery and the photovoltaic panel constraints. Through case studies, It is shown that the coordination between the photovoltaic panel and the battery is effective to reduce the total electricity cost, the discomfort and the standard deviation of power consumed especially in summer conditions [17][18].

7.3. Fractal based-decomposition optimization algorithm

Participants: El-Ghazali Talbi (externel collaborators: Amir Nakib - Univ. Paris 12 France)

In this work a new metaheuristic based on geometric fractal decomposition to solve large-scale continuous optimization problems is proposed. It consists of dividing the feasible search space into sub-regions with the same geometrical pattern. At each iteration, the most promising ones are selected and further decomposed. This approach tends to provide a dense set of samples and has interesting theoretical convergence properties. Under some assumptions, this approach covers all the search space only in case of small dimensionality problems. The aim of this work is to propose a new algorithm based on this approach with low complexity and which performs well in case of large-scale problems. To do so, a low complex method that profits from fractals properties is proposed. Then, a deterministic optimization procedure is proposed using a single solution-based metaheuristic which is exposed to illustrate the performance of this strategy. Obtained results on common test functions were compared to those of algorithms from the literature and proved the efficiency of the proposed algorithm [8].

7.4. Parallel High-Level Search Heuristics for Single- and Multi-objective Optimization

Participants: Bilel Derbel, Arnaud Liefooghe (external collaborators: Sebastien Verel, Univ. Littoral, France; Jialong Shi and Qingfu Zhang, City University, Hong Kong)

A parallel tabu search for the unconstrained binary quadratic programming problem. Although several sequential heuristics have been proposed for dealing with the Unconstrained Binary Quadratic Programming (UBQP), very little effort has been made for designing parallel algorithms for the UBQP. In [26], we propose a novel decentralized parallel search algorithm, called Parallel Elite Biased Tabu Search (PEBTS). It is based on D2TS, a state-of-the-art sequential UBQP metaheuristic. The key strategies in the PEBTS algorithm include: (i) a lazy distributed cooperation procedure to maintain diversity among different search processes and (ii) finely tuned bit-flip operators which can help the search escape local optima efficiently. Our experiments on the Tianhe-2 supercomputer with up to 24 computing cores show the accuracy of the efficiency of PEBTS compared with a straightforward parallel algorithm running multiple independent and non-cooperating D2TS processes.

Decomposition-based parallel strategies to speed up Pareto local search. Pareto Local Search (PLS) is a basic building block in many state-of-the-art multiobjective combinatorial optimization algorithms. However, the basic PLS requires a long time to find high-quality solutions. In [27], we propose and investigate several parallel strategies to speed up PLS using decomposition. These strategies are based on a parallel multi-search framework. In our experiments, we investigate the performances of different parallel variants of PLS on the multiobjective unconstrained binary quadratic programming problem. Each PLS variant is a combination of the proposed parallel strategies. The experimental results show that the proposed approaches can significantly speed up PLS while maintaining about the same solution quality. In addition, we introduce a new way to visualize the search process of PLS on two-objective problems, which is helpful to understand the behaviors of PLS algorithms.

7.5. Large scale GPU-centric optimization

Participants: J. Gmys, M. Gobert and N. Melab
This contribution is a joint work with M. Mezmaz and D. Tuyttens from University of Mons (UMONS), and T. C. Pessoa and F. H. De Carvalho Junior from Universidade Federal Do Ceará (UFC), Brazil. N. Melab and M. Mezmaz have been the guest editors [7] of a special issue in the CCPE journal on this topic.
Nowadays, accelerator-centric architectures offer orders-of-magnitude performance and energy improvements. The interest of those parallel resources has been recently accentuated by the advent of deep learning making them definitely key-building blocks of modern supercomputers. During the year 2017, the focus has been put on the investigation of these specific devices within the context of parallel optimization. In the following, two major contributions are reported: (1) massively parallel GPU-centric tree-based optimization for solving to optimality big permutation optimization problems; (2) Cuda Dynamic Parallelism (CDP) for backtracking. Another contribution [2] on the parallel solving of permutation (flow-shop) problems is proposed but not presented here.

- **Massively parallel GPU-centric tree-based optimization.** Within the context of the Ph.D thesis (jointly supervised with UMONS, Belgium) of Jan Gmys [1], parallel Branch-and-Bound (B&B) has been revisited for multi-core processors, (multi-)GPU accelerators and MIC coprocessors [6]. During the 2017 year, the focus was put on the extension of these contributions in order to deal with large hybrid clusters. A bi-level parallel approach is actually proposed to revisit the parallel B&B at the intra- and inter-processing node levels. The intra-node level consists in the combination of the previous contributions for an efficient exploitation of the parallelism levels provided inside a processing node which can be a multi-core processor, a GPU, a Xeon Phi or their combination (hybrid processing node). The inter-node (cluster) level deals with the parallelism provided through the use of multiple processing nodes. In the PhD thesis of M. Mezmaz, we have investigated such parallelism by proposing a grid-enabled approach called B&B@Grid, including interval-based work stealing (WS) and checkpointing mechanisms. In addition, each processing node is mainly composed of a single processing core processing a single interval at a time. Our contribution consists in revisiting those mechanisms to deal with multi- and many-core resources. Indeed, for instance a GPU explores thousands of intervals. The contribution includes a bi-level WS mechanism together with a multi-interval checkpointing mechanism. The proposed approach has been experimented on the OUESSANT GPU cluster located at IDRIS, Paris. The results show that, on the road to the exascale era, our approach scales up to 130.000 Cuda cores, reducing the execution time from 25 days (using B&B@Grid) to 9 hours.
- **Cuda Dynamic Parallelism (CDP) for backtracking.** New GPGPU technologies, such as CUDA Dynamic Parallelism (CDP), can help dealing with recursive patterns of computation, such as divide-and-conquer, used by backtracking algorithms. During 2017, in collaboration with University of Cearà (Brazil), we have investigated the CDP mechanism using highly irregular algorithms. Indeed, we have proposed a GPU-accelerated backtracking algorithm using CDP that extending a well-known parallel backtracking model. The algorithm analyzes the memory requirements of subsequent kernel generations and performs no dynamic memory allocation on GPU, unlike related works from the literature. The proposed algorithm has been extensively experimented using the N-Queens Puzzle

problem and instances of the Asymmetric Traveling Salesman Problem (ATSP) as test-cases. The proposed CDP algorithm may, under some conditions, outperform its non-CDP counterpart by a factor up to 25. Compared to other CDP-based strategies from the literature, the proposed algorithm is on average $8\times$ faster.

INOCS Team

7. New Results

7.1. Large scale complex structure optimization

New decomposition methods for the time-dependent combined network design and routing problem: A significant amount of work has been focussed on the design of telecommunication networks. The performance of different Integer Programming models for various situations has been computationally assessed. One of the settings that has been thoroughly analyzed is a variant where routing decisions (for time-dependent traffic demand), and network design, are combined in a single optimization model. Solving this model with a state-of-the-art solver on representative network topologies, shows that this model quickly becomes intractable. With an extended formulation, both the number of continuous flow variables and the number of fixed charge capacity constraints are multiplied by a factor $|V|$ (where V represents the set of nodes) leading to large model. However, the linear relaxation of this extended formulation yields much better lower bounds. Nevertheless, even if the extended model provides stronger lower bounds than the aggregated formulation, it suffers from its huge size: solving the linear relaxation of the problem quickly becomes intractable when the network size increases, making the linear relaxation expensive to solve. This observation motivates the analysis of decomposition methods [21].

Convex piecewise linear unsplittable multicommodity flow problems We studied the multi-commodity flow problem with unsplittable flows, and piecewise-linear costs on the arcs. They show that this problem is NP-hard when there is more than one commodity. We propose a new MILP models for this problem, that was compared to two formulations commonly used in the literature. The computational experiments reveal that the new model is able to obtain very strong lower bounds, and is very efficient to solve the considered problem [22].

Tree Reconstruction Problems: We studied the problem of reconstructing a tree network by knowing only its set of terminal nodes and their pairwise distances, so that the reconstructed network has its total edge weight minimized. This problem has applications in several areas, namely the inference of phylogenetic trees and the inference of routing networks topology. Phylogenetic trees allow the understanding of the evolutionary history of species and can assist in the development of vaccines and the study of biodiversity. The knowledge of the routing network topology is the basis for network tomography algorithms and it is a key strategy to the development of more sophisticated and ambitious traffic control protocols and dynamic routing algorithms [24].

Distribution network configuration problems: A distribution network is a system aiming to transfer a certain type of resource from feeders to customers. Feeders are producers of a resource and customers have a certain demand in this resource that must be satisfied. Distribution networks can be represented on graphs and be subject to constraints that limit the number of intermediate nodes between some elements of the network (hop constraints) because of physical constraints. We used layered graphs for hop constrained problems to build extended formulations [16]. Preprocessing techniques allowed to reduce the size of the layered graphs used. The model was studied on the hop-constrained minimum margin problem in an electricity network. This problem consists of designing a connected electricity distribution network, and to assign customers to electricity feeders at a maximum number of hops so as to maximize the minimum capacity margin over the feeders to avoid an overload for any feeder.

Comparison of formulations and solution methods for location problems: We addressed two classes of location problems the Discrete Ordered Median Problem (DOMP) and the two-level uncapacitated facility location problem with single assignment constraints (TUFLP-S), an extension of the uncapacitated facility location problem. We presented several new formulations for the DOMP based on its similarity with some scheduling problems. Some of the new formulations present a considerably smaller number of constraints

to define the problem with respect to some previously known formulations. Furthermore, the lower bounds provided by their linear relaxations improve the ones obtained with previous formulations in the literature even when strengthening is not applied. We also present a polyhedral study of the assignment polytope of our tightest formulation showing its proximity to the convex hull of the integer solutions of the problem. Several resolution approaches, among which we mention a branch and cut algorithm, are compared. Extensive computational results on two families of instances, namely randomly generated and from Beasley's OR-library, show the power of our methods for solving DOMP [28]. We also addressed the TUFLP-S for which we presented six mixed-integer programming models based on reformulation techniques and on the relaxation of the integrality of some of the variables associated with location decisions. We compared the models by carrying out extensive computational experiments on large, hard, artificial instances, as well as on instances derived from an industrial application in freight transportation [27].

New models and algorithms for integrated vehicle routing problems: We address a real-life inventory routing problem, which consists in designing routes and managing the inventories of the customers simultaneously. The problem was introduced during the 2016 ROADEF/EURO challenge. The proposed problem is original and complex for several reasons: the logistic ratio optimization objective, the hourly time-granularity for inventory constraints over a large planning horizon, the driver/trailer allocation management. Clearly, this problem is an optimization problem with complex structure, for which we propose an extended formulation that we address with a heuristic branch-price-and-cut method. Among the difficulties, that we had to face, are: the fractional objective function, the simultaneous generation of constraints and columns, and a complex pricing problem. We evaluate our approach on the benchmark instances proposed by the enterprise Air Liquide co-organiser of challenge. The solution method allowed the team including INOCS members to win the scientific prize of the ROADEF/EURO challenge 2016 [47]. We also addressed a rich Traveling Salesman Problem with Profits encountered in several real-life cases. We proposed a unified solution approach based on variable neighborhood search. Our approach includes two loading neighborhoods based on the solution of mathematical programs are proposed to intensify the search. They interact with the routing neighborhoods as it is commonly done in matheuristics. The performance of the proposed matheuristic is assessed on various instances proposed for the Orienteering Problem and the Orienteering Problem with Time Window including up to 288 customers. The computational results show that the proposed matheuristic is very competitive compared with the state-of-the-art methods. Extensive computational experiments on the new testbed confirm the efficiency of the matheuristic [30].

A heuristic approach to solve an integrated warehouse order picking problem: We study an integrated warehouse order picking problem with manual picking operations. The picking area of the warehouse is composed by a set of storage positions. The working day is divided in periods. For each period, each position contains several pieces of a unique product, and a set of customer orders has to be prepared. An order is a set of products, each associated with a quantity. The order pickers can prepare up to K parcels in a given picking route. The problem consists in jointly deciding: (1) the assignment of references to storage positions in the aisles which need to be filled up; (2) the division of orders into several parcels, respecting weight and size constraints; (3) the batching of parcels into groups of size K, that implicitly define the routing into the picking area. The objective function is to minimize the total walking distance. In order to deal with industrial instances of large size (considering hundreds of clients, thousands of positions and product references) in a short computation time, a heuristic method based on dynamic programming and minimum cost flow paradigms is proposed. Experimental results on an industrial benchmark have reported very good results with respect to the actual industrial solution [64].

New models for Load Scheduling for Residential Demand Response on Smart Grids. The residential load scheduling problem is concerned with finding an optimal schedule for the operation of residential loads so as to minimize the total cost of energy while aiming to respect a prescribed limit on the power level of the residence. We propose a mixed integer linear programming formulation of this problem that accounts for the consumption of appliances, generation from a photovoltaic system, and the availability of energy storage. A distinctive feature of our model is the use of operational patterns that capture the individual operational characteristics of each load, giving the model the capability to accommodate a wide range of possible operating

patterns for the flexible loads. The proposed formulation optimizes the choice of operational pattern for each load over a given planning horizon. In this way, it generates a schedule that is optimal for a given planning horizon, unlike many alternatives based on controllers. The formulation can be incorporated into a variety of demand response systems, in particular because it can account for different aspects of the cost of energy, such as the cost of power capacity violations, to reflect the needs or requirements of the grid. Our computational results show that the proposed formulation is able to achieve electricity costs savings and to reduce peaks in the power consumption, by shifting the demand and by efficiently using a battery [52].

Lagrangian heuristics for SVM with feature selection: The focus of pattern classification is to recognize similarities in the data, categorizing them in different subsets. In many fields, such as the financial and the medical ones, classification of data (samples in Machine Learning language) is useful for analysis or diagnosis purposes. Quite often datasets are formed by a small number of samples, which in turn are characterized by a huge number of attributes (features). The handling of the entire feature set would be computationally very expensive and its outcome would lack from insight. For this reason, it is convenient to reduce the set of features which is expected to be easier to interpret and also easy to evaluate. However, it is not always easy to predict which of those are relevant for classification purposes. Hence it is necessary to screen off the relevant features from those which are irrelevant. The process that selects the features entering the subset of the relevant ones is known as Feature Selection (FS). The (FS) problem can be treated explicitly as a Mixed Binary Programming (MBP) one in the framework of the Support Vector Machine approach. We have discussed a Lagrangian-relaxation-based heuristics. In particular we embed into our objective function a weighted combination of the L^1 and L^0 norm of the normal to the separating hyperplane. We come out with a Mixed Binary Linear Programming problem which is suitable for a Lagrangian relaxation approach. Based on a property of the optimal multiplier setting, we apply a consolidated nonsmooth optimization ascent algorithm to solve the resulting Lagrangian dual. In the proposed approach we get, at every ascent step, both a lower bound on the optimal solution as well as a feasible solution at low computational cost [26].

Decomposition methods for tree-based network design problems: We studied different problems, where the underlying solution structure needs to have a tree-like topology and some additional constraints need to be fulfilled. For all these problems, we focused on solution approaches, which allow to tackle large-scale instances, as the application of these problems in areas like systems biology often has to deal with instances containing tens of thousands of nodes. In order to solve these problems efficiently, we turned to decomposition methods, like Benders decomposition, Lagrangian relaxation or relax-and-cut. The considered problems include the (prize-collecting) Steiner tree problem [17], [33], tree-star problems [32], the shared arborescence problem [34], the upgrading spanning tree problem [36] and for maximum-weight connected subgraph problems [35].

We also studied models arising in the design of switched Ethernet networks implementing the Multiple Spanning Tree Protocol [24]. In these problems, multiple spanning trees have to be established in a network to route demands partitioned into virtual local access networks. Different mixed-integer formulations for the problem have been proposed and compared, both theoretically and computationally.

Dynamic programming for the minimum-cost maximal knapsack packing problem: Given a set of items with profits and weights and a knapsack capacity, we studied the problem of finding a maximal knapsack packing that minimizes the profit of the selected items. We proposed an effective dynamic programming (DP) algorithm which has a pseudo-polynomial time complexity. We demonstrated the equivalence between this problem and the problem of finding a minimal knapsack cover that maximizes the profit of the selected items. In an extensive computational study on a large and diverse set of benchmark instances, we demonstrated that the new DP algorithm outperforms a state-of-the-art commercial mixed-integer programming (MIP) solver applied to the two best performing MIP models from the literature [25].

7.2. Bilevel Programming

Bilevel approaches for energy management problems: We have proposed the first bilevel pricing models to explore the relationship between energy suppliers and customers who are connected to a smart grid. Due to

their definition, bilevel models enable to integrate customer response into the optimization process of supplier who aims to maximize revenue or minimize capacity requirements. In our setting, the energy provider acts as a leader (upper level) that takes into account a smart grid (lower level) that minimizes the sum of users' disutilities. The latter bases its decisions on the hourly prices set by the leader, as well as the schedule preferences set by the users for each task. Moreover the follower is able to produce renewable energy and store it. The pricing problems, we model, belong to the category of stochastic single leader single follower problems. A scenario based approach is based to solve the problem. For each scenario, the bilevel program is solved by rewriting it as a single level optimization program. Numerical results on randomly generated instances illustrate numerically the validity of the approach, which achieves an optimal trade-off between three objectives: revenue, user cost, and peak demand [53].

Network pricing problems with unit toll: In the so-called network pricing problem an authority owns some arcs of the network and tolls them, while users travel between their origin and destination choosing their minimum cost path. We consider a unit toll scheme, and in particular the cases where the authority is imposing either the same toll on all of its arcs, or a toll proportional to a given parameter particular to each arc (for instance a per kilometer toll). We show that if tolls are all equal then the complexity of the problem is polynomial, whereas in case of proportional tolls it is pseudo-polynomial, proposing ad-hoc solution algorithms and relating these problems to the parametric shortest path problem. We then address a robust approach using an interval representation to take into consideration uncertainty on parameters. We show how to modify the algorithms for the deterministic case to solve the robust counterparts, maintaining their complexity class [15].

New formulations for solving Stackelberg games: We analyzed general Stackelberg games (SGs) and Stackelberg security games (SSGs). SGs are hierarchical adversarial games where players select actions or strategies to optimize their payoffs in a sequential manner. SSGs are a type of SGs that arise in security applications, where the strategies of the player that acts first consist in protecting subsets of targets and the strategies of the followers consist in attacking one of the targets. We review existing mixed integer optimization formulations in both the general and the security setting and present new formulations for the the second one. We compare the SG formulations and the SSG formulations both from a theoretical and a computational point of view. We identify which formulations provide tighter linear relaxations and show that the strongest formulation for the security version is ideal in the case of one single attacker. Our computational experiments show that the new formulations can be solved in shorter times [66].

A branch and price algorithm for solving Stackelberg Security games: Mixed integer optimization formulations are an attractive alternative to solve Stackelberg Game problems thanks to the efficiency of state of the art mixed integer algorithms. In particular, decomposition algorithms, such as branch and price methods, make it possible to tackle instances large enough to represent games inspired in real world domains. We focus on Stackelberg Games that arise from a security application and investigate the use of a new branch and price method to solve its mixed integer optimization formulation. We prove that the algorithm provides upper and lower bounds on the optimal solution at every iteration and investigate the use of stabilization heuristics. Our preliminary computational results compare this solution approach with previous decomposition methods obtained from alternative integer programming formulations of Stackelberg games [29].

A new general-purpose algorithm for mixed-integer bilevel linear programs: We considered bilevel problems with continuous and discrete variables at both levels, with linear objectives and constraints (continuous upper level variables, if any, must not appear in the lower level problem). We proposed a general-purpose branch-and-cut exact solution method based on several new classes of valid inequalities, which also exploits a very effective bilevel-specific preprocessing procedure. An extensive computational study was presented to evaluate the performance of various solution methods on a common testbed of more than 800 instances from the literature and 60 randomly generated instances. Our new algorithm consistently outperformed (often by a large margin) alternative state-of-the-art methods from the literature, including methods exploiting problem-specific information for special instance classes. In particular, it solved to optimality more than 300

previously unsolved instances from the literature. To foster research on this challenging topic, our solver was made publicly available online [18], [19].

A mixed-integer programming based heuristic for generalized interdiction problems: We considered a subfamily of mixed-integer linear bilevel problems that we call Generalized Interdiction Problems. This class of problems includes, among others, the widely-studied interdiction problems, i.e., zero-sum Stackelberg games where two players (called the leader and the follower) share a set of items, and the leader can interdict the usage of certain items by the follower. Problems of this type can be modeled as Mixed-Integer Nonlinear Programming problems, whose exact solution can be very hard. We propose a new heuristic scheme based on a single-level and compact mixed-integer linear programming reformulation of the problem obtained by relaxing the integrality of the follower variables. A distinguished feature of our method is that general-purpose mixed-integer cutting planes for the follower problem are exploited, on the fly, to dynamically improve the reformulation. The resulting heuristic algorithm proved very effective on a large number of test instances, often providing an (almost) optimal solution within very short computing times. [20]

Unit Commitment under Market Equilibrium Constraints: Traditional (deterministic) models for the Unit Commitment problem (UC) assume that the net demand for each period is perfectly known in advance, or in more recent and more realistic approaches, that a set of possible demand scenarios is known (leading to stochastic or robust optimization problems). However, in practice, the demand is dictated by the amounts that can be sold by the producer at given prices on the day-ahead market. We modeled and solved the UC problem with a second level of decisions ensuring that the produced quantities are cleared at market equilibrium. In its simplest form, we are faced to a bilevel optimization problem where the first level is a MIP and the second level linear. As a first approach to the problem, we assumed that demand curves and offers of competitors in the market are known to the operator. Following the classical approach for these models, we turned the problem into a single-level program by rewriting and linearizing the first-order optimality conditions of the second level [50].

7.3. Robust/Stochastic programming

Decomposition method for stochastic staff management problems: We addressed an integrated shift scheduling and load assignment optimization problem for attended home delivery, which is a last-mile delivery service requiring the presence of the customer for the delivery. We were mainly interested in generating a daily master plan for each courier. We proposed a tactical problem integrating a shift scheduling problem and a load assignment problem under demand uncertainty, which was modeled as a two-stage stochastic programming model. This model integrates two types of decisions. First-stage decisions are related to the design of a schedule that includes the periods of the day in which each courier must work and the o-d pairs to visit at each time period. Second-stage decisions (recourse actions) consist of the allocation of a number of packages to be delivered at each time period, for each o-d pair, by each courier, such that the demand (number of packages to deliver) for each scenario is satisfied. Recourse is the ability to take corrective actions after a random event has taken place. The objective is to minimize the sum of the daily staffing cost plus the expected daily recourse cost. To solve this problem, we proposed and implemented a multi-cut integer L-shaped algorithm, where the second stage decomposes by time periods and by demand scenarios. To strengthen the first stage model, some valid inequalities are added, and some of the existing constraints are lifted. Moreover, we addressed the operational planning problem which aims to incorporate the tactical master plan solutions into the real-time allocation of client requests to the couriers. We proposed a mathematical model and a solution approach based on a column generation algorithm. The goal of this approach was to provide a tool to evaluate the robustness of the tactical plan, i.e. how well this plan reacts to new order requests arriving in real-time. Results on real-world based instances from a delivery company demonstrate that our approach provides robust tactical solutions that easily accommodate to fluctuations in customer orders, preventing additional costs related to the underutilization of couriers and the use of external couriers to satisfy all delivery requests [38], [65].

Decomposition method for the stochastic Steiner tree problem: We introduced a new algorithmic approach for solving the stochastic Steiner tree problem based on three procedures for computing lower bounds (dual ascent, Lagrangian relaxation, Benders decomposition). Our method is derived from a new integer linear

programming formulation, which is shown to be strongest among all known formulations. The resulting method, which relies on an interplay of the dual information retrieved from the respective dual procedures, computes upper and lower bounds and combines them with several rules for fixing variables in order to decrease the size of problem instances. The effectiveness of our method is compared in an extensive computational study with the state-of-the-art exact approach, which employs a Benders decomposition based on two-stage branch-and-cut, and a genetic algorithm introduced during the DIMACS implementation challenge on Steiner trees. Our results indicate that the presented method significantly outperforms existing ones, both on benchmark instances from literature, as well as on large-scale telecommunication networks [31].

MEPHYSTO Project-Team

4. New Results

4.1. Long-time homogenization of the wave equation

In a joint work [36], A. Benoit and A. Gloria considered an elliptic operator in divergence form with symmetric coefficients. If the diffusion coefficients are periodic, the Bloch theorem allows one to diagonalize the elliptic operator, which is key to the spectral properties of the elliptic operator and the usual starting point for the study of its long-time homogenization. When the coefficients are not periodic (say, quasi-periodic, almost periodic, or random with decaying correlations at infinity), the Bloch theorem does not hold and both the spectral properties and the long-time behavior of the associated operator are unclear. At low frequencies, we may however consider a formal Taylor expansion of Bloch waves (whether they exist or not) based on correctors in elliptic homogenization. The associated Taylor-Bloch waves diagonalize the elliptic operator up to an error term (an "eigendefect"), which we express with the help of a new family of extended correctors. We use the Taylor-Bloch waves with eigendefects to quantify the transport properties and homogenization error over large times for the wave equation in terms of the spatial growth of these extended correctors. On the one hand, this quantifies the validity of homogenization over large times (both for the standard homogenized equation and higher-order versions). On the other hand, this allows us to prove asymptotic ballistic transport of classical waves at low energies for almost periodic and random operators.

4.2. Weighted functional inequalities

Functional inequalities like spectral gap, covariance, or logarithmic Sobolev inequalities are powerful tools to prove nonlinear concentration of measure properties and central limit theorem scalings. Besides their well-known applications in mathematical physics (e.g. for the study of interacting particle systems like the Ising model or for interface models), such inequalities were recently used by the team to establish quantitative stochastic homogenization results.

These functional inequalities have nevertheless two main limitations for stochastic homogenization. On the one hand, whereas only few examples are known to satisfy them (besides product measures, Gaussian measures, and more general Gibbs measures with nicely behaved Hamiltonians), these inequalities are not robust with respect to various simple constructions: for instance, a Poisson point process satisfies a spectral gap, but the random field corresponding to the Voronoi tessellation of a Poisson point process does not. On the other hand, these functional inequalities require random fields to have an integrable covariance, which prevents one to consider fields with heavier tails.

In the series of work [26], [27], [28], M. Duerinckx and A. Gloria introduced weaker versions of these functional inequalities in the form of weighted inequalities. The interest of such inequalities is twofold: first, as their unweighted counterpart they ensure strong concentration properties; second, they hold for a large class of statistics of interest to homogenization (which is shown using a constructive approach).

4.3. Macroscopic behaviors of large interacting particle systems

A vast amount of physical phenomena were first described at the macroscopic scale, in terms of the classical partial differential equations (PDEs) of mathematical physics. Over the last decades the scientific community has pursued part of its research towards the following *universality principle*, which is well known in statistical physics: "the qualitative behavior of physical systems depend on the microscopic details of the system only through some large-scale variables". Typically, the microscopic systems are composed of a huge number of atoms and one looks at a very large time scale with respect to the typical frequency of atom vibrations. Mathematically, this corresponds to a space-time scaling limit procedure.

The macroscopic laws that can arise from microscopic systems can either be partial differential equations (PDEs) or stochastic PDEs (SPDEs) depending on whether one is looking at the convergence to the mean or at the fluctuations around that mean. Therefore, it is a natural problem in the field of interacting particle systems to obtain the macroscopic laws of the relevant thermodynamical quantities, using an underlying microscopic dynamics, namely particles that move according to some prescribed stochastic law. Probabilistically speaking, these systems are continuous time Markov processes.

4.3.1. Anomalous diffusion

First, one can imagine that at the *microscopic* scale, the population is well modeled by stochastic differential equations (SDEs). Then, the *macroscopic* description of the population densities is provided by partial differential equations (PDEs), which can be of different types. All these systems may characterize the collective behavior of individuals in biology models, but also agents in economics and finance. In [14] M. Simon in collaboration with C. Olivera has obtained a limit process which belongs to the family of non-local PDEs, and is related to anomalous diffusions. More precisely, they study the asymptotic behavior of a system of particles which interact *moderately*, i.e. an intermediate situation between weak and strong interaction, and which are submitted to random scattering. They prove a law of large numbers for the empirical density process, which in the macroscopic limit follows a fractional conservation law. The latter is a generalization of convection-diffusion equations, and can appear in physical models (e.g. over-driven detonation in gases [38], or semiconductor growth [55]), but also in areas like hydrodynamics and molecular biology.

Another approach which aims at understanding this abnormally diffusive phenomena is to start from deterministic system of Newtonian particles, and then perturb this system with a stochastic component which will provide enough ergodicity to the dynamics. It is already well known that these stochastic chains model correctly the behavior of the conductivity [35]. In two published papers [18][32], and another submitted one [19], M. Simon with her coauthors C. Bernardin, P. Gonçalves, M. Jara, T. Komorowski, S. Olla and M. Sasada have observed both behaviors, normal and anomalous diffusion, in the context of low dimensional asymmetric systems. They manage to describe the microscopic phenomena at play which are responsible for each one of these phenomena, and they go beyond the predictions that have recently been done in [51], [52].

4.3.2. Towards the weak KPZ universality conjecture

Among the classical SPDEs is the Kardar-Parisi-Zhang (KPZ) equation which has been first introduced more than thirty years ago in [46] as the *universal* law describing the fluctuations of randomly growing interfaces of one-dimensional stochastic dynamics close to a stationary state (as for example, models of bacterial growth, or fire propagation). In particular, the *weak KPZ universality conjecture* [52] states that the fluctuations of a large class of one-dimensional microscopic interface growth models are ruled at the macroscopic scale by solutions of the *KPZ equation*. Thanks to the recent result of M. Jara and P. Gonçalves [45], one has now all in hands to establish that conjecture. In their paper, the authors introduce a new tool, called the second order Boltzmann-Gibbs principle, which permits to replace certain additive functionals of the dynamics by similar functionals given in terms of the density of the particles. In [13], M. Simon in collaboration with P. Gonçalves and M. Jara give a new proof of that principle, which does not impose the knowledge on the spectral gap inequality for the underlying model and relies on a proper decomposition of the antisymmetric part of the current of the system in terms of polynomial functions. In addition, they fully derive the convergence of the equilibrium fluctuations towards (1) a trivial process in case of super-diffusive systems, (2) an Ornstein-Uhlenbeck process or the unique *energy solution* of the stochastic Burgers equation (SBE) (and its companion, the KPZ equation), in case of weakly asymmetric diffusive systems. Examples and applications are presented for weakly and partial asymmetric exclusion processes, weakly asymmetric speed change exclusion processes and Hamiltonian systems with exponential interactions.

In [30], M. Simon together with P. Gonçalves and N. Perkowski go beyond the weak KPZ universality conjecture to derive a new SPDE, namely, the KPZ equation with boundary conditions, from an interacting particle system in contact with stochastic reservoirs. They legitimate the choice done at the macroscopic level for the KPZ/SBE equation from the microscopic description of the system. For that purpose, they prove two main theorems: first, they extend the notion of energy solutions to the stochastic Burgers equation

by adding Dirichlet boundary conditions. Second, they construct a microscopic model (based on weakly asymmetric exclusion processes) from which the energy solution naturally emerges as the macroscopic limit of its stationary density fluctuations. This gives a physical justification for the Dirichlet boundary conditions the SBE equation. They also prove existence and uniqueness of energy solutions to two related SPDEs: the KPZ equation with Neumann boundary conditions and the SHE with Robin boundary conditions, and they rigorously establish the formal links between the equations. This is more subtle than expected, because the boundary conditions do not behave canonically. Finally, they associate an interface growth model to the microscopic model, roughly speaking by integrating it in the space variable, and show that it converges to the energy solution of the KPZ equation, thereby giving a physical justification of the Neumann boundary conditions.

4.4. High order exponential integrators for nonlinear Schrödinger equations with application to rotating Bose–Einstein condensates

In a recent work with C. Besse and I. Violet [6], Guillaume Dujardin has proposed and analyzed new methods for the time integration of the nonlinear Schrödinger equation in the context of rotating Bose–Einstein condensates. In particular, he has proposed a systematic way to design high-order in time implicit exponential methods, given sufficient conditions to ensure mass preservation by the methods and proved high order in several physically relevant situations. He has compared those methods to several other popular methods from the literature and provided several numerical experiments.

4.5. Periodic modulations controlling Kuznetsov-Ma soliton formation in nonlinear Schrödinger equations

Together with colleages from the Physics department of the Université de Lille, S. de Bièvre and G. Dujardin have analyzed the exact Kuznetsov–Ma soliton solution of the one-dimensional nonlinear Schrödinger equation in the presence of periodic modulations satisfying an integrability condition [15]. They showed that, in contrast to the case without modulation, the Kuznetsov–Ma soliton develops multiple compression points whose number, shape and position are controlled both by the intensity of the modulation and by its frequency. In addition, when this modulation frequency is a rational multiple of the natural frequency of the Kutzetsov–Ma soliton, a scenario similar to a nonlinear resonance is obtained: in this case the spatial oscillations of the Kuznetsov–Ma soliton's intensity are periodic. When the ratio of the two frequencies is irrational, the soliton's intensity is a quasiperiodic function. A striking and important result of this analysis is the possibility to suppress any component of the output spectrum of the Kuznetsov–Ma soliton by a judicious choice of the amplitude and frequency of the modulation.

4.6. Exponential integrators for nonlinear Schrödinger equations with white noise dispersion

Together with D. Cohen, G. Dujardin has proposed several exponential numerical methods for the time integration of the nonlinear Schrödinger equation with power law nonlinearity and random dispersion [11]. In particular, he introduced a new explicit exponential integrator for this purpose that integrates the noisy part of the equation exactly. He prove that this scheme is of mean-square order 1 and he drew consequences of this fact. He compared the exponential integrator with several other numerical methods from the literature. Finally, he proposed a second exponential integrator, which is implicit and symmetric and, in contrast to the first one, preserves the L^2-norm of the solution.

4.7. New results on waveguides with mixed diffusion

In [21], [8], [22], D. Bonheure, J.-B. Casteras and collaborators obtained new results on the existence and qualitative properties of waveguides for a mixed-diffusion NLS equation. In particular, they proved the first existence results for waveguides with fixed mass and provided several qualitative descriptions of these. They also showed that the ground-state solutions are instable by finite (or infinite) time blow-up improving a recent result of Boulenger and Lenzmann and answering a conjecture of Baruch and Fibich.

4.8. New result on the Boltzmann scenario

Boltzmann provided a scenario to explain why individual macroscopic systems inevitably approach a unique macroscopic state of thermodynamic equilibrium, and why after having done so, they remain in that state, apparently forever. In [12], new rigorous results are provided that mathematically prove the basic features of Boltzmann's scenario for two classical models: a simple boundary-free model for the spatial homogenization of a non-interacting gas of point particles, and the well-known Kac ring model.

4.9. Other new results

In [9], [20], D. Bonheure, J.-B. Casteras and collaborators made bifurcation analysis and constructed multi-layer solutions of the Lin-Ni-Takagi and Keller-Segel equations, which come from the Keller-Segel system of chemotaxis in specific cases. 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 these models.

In [10], [23], J.-B. Casteras and collaborators study different problems related to the existence of A-harmonic functions with prescribed asymptotic boundary on Cartan-Hadamard manifold. In particular, they obtained a sharp lower bound on the section curvature for the existence of minimal graphic functions with prescribed asymptotic boundary.

In [25], a kinetic equation of the Vlasov-Wave type is studied, which arises in the description of the behavior of a large number of particles interacting weakly with an environment. Variational techniques are used to establish the existence of large families of stationary states for this system, and analyze their stability.

MODAL Project-Team

7. New Results

7.1. An Oracle Inequality for Quasi-Bayesian Non-Negative Matrix Factorization

Participant: Benjamin Guedj.

The quasi-Bayesian perspective has been extended to the popular setting of non-negative matrix factorization. This is a pivotal problem in machine learning (image segmentation, recommendation systems, audio source separation, ...) and an original estimator of the unobserved matrix has been proposed. An oracle inequality is derived, along with several possible implementations. This work is published in Mathematical Methods of Statistics [12].

It a joint work with Pierre Alquier from ENSAE - Université Paris-Saclay.

7.2. Simpler PAC-Bayesian Bounds for Hostile Data

Participant: Benjamin Guedj.

An original and much simpler way of deriving PAC-Bayesian bounds has been introduced through the use of f-divergences (therefore generalizing earlier works on Renyi's divergence and Kullback-Leibler divergence). This work is published in Machine Learning [13].

It a joint work with Pierre Alquier from ENSAE - Université Paris-Saclay.

7.3. Highlight 1 High-dimensional Adaptive Ranking with PAC-Bayesian Bounds

Participant: Benjamin Guedj.

The quasi-Bayesian perspective has been extended to the popular setting high-dimensional ranking. This is a pivotal problem in machine learning and is at the core of several applications in industry (recommender systems, active learning, ...). An original estimator of the scoring function is proposed, and we have shown its minimax optimal properties. Our procedure is adaptive to the unknown sparsity level of the data. This work is published in Journal of Statistical Planning and Inference.

It a joint work with Sylvain Robbiano from University College London.

7.4. Online Adaptive Clustering

Participant: Benjamin Guedj.

The quasi-Bayesian perspective has been extended to online adaptive clustering. Data streams are clustered dynamically with a quasi-Bayesian-flavored predictor, and we have proven minimax regret bounds. An efficient MCMC-based implementation is proposed.

7.5. Study of Transcriptional Regulation

Participant: Guillemette Marot.

The implementation of a mixture model of normal and exponential laws enabled to define a threshold on the number of co-recruiting transcriptional regulators in order to classify cis-regulatory modules. The new findings in Biology have been published in [16].

7.6. Functional Binary Linear Models for Stratified Samples

Participant: Sophie Dabo-Niang.

Sophie Dabo-Niang's new result concern a work on functional binary linear models for stratified samples. This work introduces a new functional binary choice model in a case-control or choice-based sample design context, where the response is binary, while the explanatory variable is functional. The model is estimated when the sample is stratified with respect to the values of the response variable. A dimensional reduction of the space of the explanatory random function based on a Karhunen-Loève expansion is used to define a conditional maximum likelihood estimate of the model. Based on this formulation, several asymptotic properties are given. Numerical experiments are used to compare the proposed method with the ordinary maximum likelihood method, which ignores the nature of the sampling. The proposed model yields encouraging results.

7.7. Mixture Model for Mixed Kind of Data

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

A mixture model of Gaussian copula allows to cluster mixed kind of data has been proposed. Each component is composed by classical margins while the conditional dependencies between the variables is modeled by a Gaussian copula. The parameter estimation is performed by a Gibbs sampler. This work has been now published to an international journal [18]. Furthermore, an R package (MixCluster) is available on Rforge.

7.8. Data Units Selection in Statistics

Participant: Christophe Biernacki.

Usually, the data unit definition is fixed by the practitioner but it can happen that he/her hesitates between several data unit options. In this context, it is highlighted that it is possible to embed data unit selection into a classical model selection principle. The problem is introduced in a regression context before to focus on the model-based clustering and co-clustering context, for data of different kinds (continuous, count, categorical). This work is now in revision for an international journal [36]. It has led also to three invitations as a plenary session speaker to international or national conferences (the US Classification Society Conference [27], the French Classification Society Conference [28], the StatLearn conference [20]).

It is a joint work with Alexandre Lourme from University of Bordeaux.

7.9. Trade-off Between Computation Time and Accuracy

Participants: Christophe Biernacki, Maxime Brunin, Alain Célisse.

Most estimates practically arise from algorithmic processes aiming at optimizing some standard, but usually only asymptotically relevant, criteria. Thus, the quality of the resulting estimate is a function of both the iteration number and also the involved sample size. An important question is to design accurate estimates while saving computation time, and we address it in the simplified context of linear regression here. Fixing the sample size, we focus on estimating an early stopping time of a gradient descent estimation process aiming at maximizing the likelihood. It appears that the accuracy gain of such a stopping time increases with the number of covariates, indicating potential interest of the method in real situations involving many covariates. This work has been presented to an international conference [20] and a national conference [29], and a preprint is still being in progress.

Maxime Brunin will defend his PhD thesis related to this topic on January 2018.

7.10. Projection Under Pairwise Control

Participant: Christophe Biernacki.

Visualization of high-dimensional and possibly complex (non continuous for instance) data onto a low-dimensional space may be difficult. Several projection methods have been already proposed for displaying such high-dimensional structures on a lower-dimensional space, but the information lost is not always easy to use. Here, a new projection paradigm is presented to describe a non-linear projection method that takes into account the projection quality of each projected point in the reduced space, this quality being directly available in the same scale as this reduced space. More specifically, this novel method allows a straightforward visualization data in $\mathbb{R}^2$ with a simple reading of the approximation quality, and provides then a novel variant of dimensionality reduction.

This work is still under revision in an international journal [39].

It is a joint work with Hiba Alawieh and Nicolas Wicker, both from University of Lille.

7.11. Real-time Audio Sources Classification

Participants: Christophe Biernacki, Maxime Baelde.

Recent research on machine learning focuses on audio source identification in complex environments. They rely on extracting features from audio signals and use machine learning techniques to model the sound classes. However, such techniques are often not optimized for a real-time implementation and in multi-source conditions. It is proposed here a new real-time audio single-source classification method based on a dictionary of sound models (that can be extended to a multi-source setting). The sound spectrums are modeled with mixture models and form a dictionary. The classification is based on a comparison with all the elements of the dictionary by computing likelihoods and the best match is used as a result. It is found that this technique outperforms classic methods within a temporal horizon of 0.5s per decision (achieved 6errors on a database composed of 50 classes). This work has been now extended with success to the multi-sources classification case and also the computational load has been sufficiently reduced to reach the real time target (less than 50ms). This work has been presented to an international conference in Signal Processing [25] and also to a national conference [26]. A preprint is well advanced and should be submitted to an international journal at the end of 2017.

It is a joint work with Raphaël Greff, from the A-Volute company.

7.12. Model-Based Co-clustering for Ordinal Data

Participant: Christophe Biernacki.

A model-based co-clustering algorithm for ordinal data is presented. This algorithm relies on the latent block model embedding a probability distribution specific to ordinal data (the so-called BOS or Binary Ordinal Search distribution). Model inference relies on a Stochastic EM algorithm coupled with a Gibbs sampler, and the ICL-BIC criterion is used for selecting the number of co-clusters (or blocks). The main advantage of this ordinal dedicated co-clustering model is its parsimony, the interpretability of the co-cluster parameters (mode, precision) and the possibility to take into account missing data. Numerical experiments on simulated data show the efficiency of the inference strategy, and real data analyses illustrate the interest of the proposed procedure. The resulting work is in revision to an international journal [40].

This is joint work Julien Jacques from University of Lyon 2.

7.13. Model-Based Co-clustering for Ordinal Data of different dimensions

Participant: Christophe Biernacki.

This work has been motivated by a psychological survey on women affected by a breast tumor. Patients replied at different moments of their treatment to questionnaires with answers on ordinal scale. The questions relate to aspects of their life called dimensions. To assist the psychologists in analyzing the results, it is useful to emphasize a structure in the dataset. The clustering method achieves that by creating groups of individuals that are depicted by a representative of the group. From a psychological position, it is also useful to observe how questions may be grouped. This is why a clustering should be performed also on the features, which is called a co-clustering problem. However, gathering questions that are not related to the same dimension does not make sense from a psychologist stance. Therefore, the present work corresponds to perform a constrained co-clustering method aiming to prevent questions from different dimensions from getting assembled in a same column-cluster. In addition, evolution of co-clusters along time has been investigated. The method relies on a constrained Latent Block Model embedding a probability distribution for ordinal data. Parameter estimation relies on a Stochastic EM-algorithm associated to a Gibbs sampler, and the ICL-BIC criterion is used for selecting the numbers of co-clusters. The resulting work has been submitted to an international journal [42].

This is joint work with Margot Selosse and Julien Jacques, both from University of Lyon 2. Margot Selosse is a new PhD student co-supervised by Julien Jacques and Christophe Biernacki.

7.14. MASSICCC Platform for SaaS Software Availability

Participants: Christophe Biernacki, Vincent Kubicki, Jonas Renault, Josselin Demont, Matthieu Marbac.

MASSICCC is a demonstration platform giving access through a SaaS (service as a software) concept to data analysis libraries developed at Inria. It allows to obtain results either directly through a website specific display (specific and interactive visual outputs) or through an R data object download. It started in October 2015 for two years and is common to the Modal team (Inria Lille) and the Select team (Inria Saclay). In 2016, two packages have been integrated: Mixmod and MixtComp (see the specific section about MixtComp). In 2017, the BlockCluster package has been integrated and also a particular attention to provide meaningful graphical outputs (for Mixmod, MixtComp and BlockCluster) directly in the web platform itself has led to some specific developments.

MASSICCC has led to a second short meeting in February 2017 in Lille (after a first short meeting in April 2016 in Lille) for obtaining a feedback from company and academic users.

The MASSICCC platform is available here in the web: https://massiccc.lille.inria.fr

7.15. Model-Based Co-Clustering of Multivariate Functional Data

Participant: Christophe Biernacki.

High dimensional data clustering is an increasingly interesting topic in the statistical analysis of heterogeneous large-scale data. We consider the problem of clustering heterogeneous high-dimensional data where the individuals are described by functional variables which exhibit a dynamical longitudinal structure. We address the issue in the framework of model-based co-clustering and propose the functional latent block model (FLBM). The introduced FLBM model allows to simultaneously cluster a sample of multivariate functions into a finite set of blocks, each block being an association of a cluster over individuals and a cluster over functional variables. Furthermore, the homogeneous set within each block is modeled with a dedicated latent process functional regression model which allows its segmentation according to an underlying dynamical structure. The proposed model allows thus to fully exploit the structure of the data, compared to classical latent block clustering models for continuous non functional data, which ignores the functional structure of the observations. The FLBM can therefore serve for simultaneous co-clustering and segmentation of multivariate non-stationary functions. We propose a variational expectation-maximization (EM) algorithm (VEM-FLBM) to monotonically maximize a variational approximation of the observed-data log-likelihood for the unsupervised inference of the FLBM model. This work has been presented as an invited speaker to the 61th World Staistics Congress [30].

This is a joint work with Faicel Chamroukhi of University of Caen.

7.16. Reject Inference Methods in Credit Scoring: A Rational Review

Participants: Christophe Biernacki, Adrien Ehrhardt, Vincent Vandewalle.

The granting process of all credit institutions rejects applicants having a low credit score. Developing a scorecard, *i.e.* a correspondence table between a client's characteristics and his score, requires a learning dataset in which the target variable good/bad borrower is known. Rejected applicants are *de facto* excluded from the process. This biased learning population might have deep consequences on the scorecard relevance. Some works, mostly empirical ones, try to exploit rejected applicants in the scorecard building process. This work proposes a rational criterion to evaluate the quality of a scoring model for the existing Reject Inference methods and dig out their implicit mathematical hypotheses. It is shown that, up to now, no such Reject Inference method can guarantee a better credit scorecard. These conclusions are illustrated on simulated and real data from the french branch of Crédit Agricole Consumer Finance (CACF). An early version of this work has been presented as a talk in the national conference [31] and a preprint is being to be finalized.

This is a joint work with Philippe Heinrich of University of Lille and Sébastien Beben of Crédit Agricole Consumer Finance.

7.17. Survival Analysis with Complex Covariates: A Model-based Clustering Preprocessing Step

Participants: Christophe Biernacki, Vincent Vandewalle.

Many covariates are now available through sensors in the industrial context, and are expected to be related to the survival analysis target. Such covariates are often complex, what has to be understood as a possible mix between continuous, categorical, even functional over time, variables with the possibility to contain missing or uncertain values. A natural question in survival analysis is to design in both flexible and easy way an hazard function related to these potentially complex covariates, while preserving the opportunity to benefit from classical hazard functions.

In the context of a bilateral contract with Alstom company on the survival analysis topic, we have been invited to give a tutorial in the IEEE PHM International Conference on Prognostics and Health Management in US [22] . In this tutorial, we have described how to decompose the unknown targeted hazard function into two complementary parts. The first one can be any classical user hazard function conditional on a latent categorical variable. The second one is the distribution of this latent variable conditionally to the complex covariates. The way to combine both parts is to sum their product over the latent variable (marginal distribution), leading to the final targeted hazard function. The key to perform this approach is to focus on the latent variable definition which can be obtained with a model-based clustering approach dedicated to complex covariates. Beyond a selected review of recent methodologies dedicated to clustering, we have described in depth some related software to perform previous clustering methods. Some case studies have been also provided in an industrial context. At the end of the talk the practitioner is thus able to perform such clustering method to use it finally with its own hazard function.

7.18. Dealing with Missing Data Through Mixture Models

Participants: Christophe Biernacki, Vincent Vandewalle.

Many data sets have missing values, however the majority of statistical methods need a complete dataset to work. Thus, practitioners often use imputation or multiple imputations to complete the data as a pre-processing step. Mixture models can be used to naturally deal with missing data in an integrated way depending on the purpose. Especially, they can be used to classify the data or derive estimates for the distances. This work as been presented in an international conference [21].

7.19. Review on Mixture Modeling and High-dimensional Clustering

Participant: Christophe Biernacki.

Following the Journées d'Études en Statistique on 2014 in Frejus, on the topic "model choice and model aggregation" where two lectures have been given respectively on mixture model and on high dimensional clustering, a book has been published in 2017 including two chapters related to these talks (respectively [33] and [34]).

The second chapter is a joint work with Cathy Maugis-Rabusseau of INSA Toulouse.

7.20. Dealing with Missing Not at Random Values in Model-based Clustering

Participant: Christophe Biernacki.

Missing values are current in modern data sets. In many situations, making the simplifying hypothesis that they are missing at random is not realistic. However, it is very challenging to propose sensible models which address the underlying missing process. We make such proposals specific to the clustering context, namely making the assumption that missing values are missing at random conditionally to clusters, thus leading to a quite natural not missing at random marginal model. A working paper is in progress.

It is a joint work with Julie Josse of Ecole Polytechnique and Gilles Celeux of Inria Saclay - Île de France.

7.21. Dealing with Several Cluster Variables

Participant: Vincent Vandewalle.

In model based clustering of quantitative data it is often supposed that only one clustering variable explains the heterogeneity of all the others variables. However, when variables come from different sources, it is often unrealistic to suppose that the heterogeneity of the data can only be explained by one variable. If such an assumption is made, this could lead to a high number of clusters which could be difficult to interpret. A model based multi-objective clustering is proposed, is assumes the existence of several latent clustering variables, each one explaining the heterogeneity of the data on some clustering projection. In order to estimate the parameters of the model an EM algorithm is proposed, it mainly relies on a reinterpretation of the standard factorial discriminant analysis in a probabilistic way. The obtained results are projections of the data on some principal clustering components allowing some synthetic interpretation of the principal clusters raised by the data. The behavior of the model is illustrated on simulated and real data. This work as been presented in an international conference [24].

NON-A Project-Team

7. New Results

7.1. Research axis 1: General annihilators (tools: ALG)

- Integro-differential equations and integro-differential algebras were studied in [91], presenting new opportunities in nonlinear control theory.
- Algebraic estimation in partial derivatives systems were studied in [93].
- An effective version of the algebraic parameter estimation problem has recently been initiated in [73] based on algebraic analysis (module theory, homological algebra) and computer algebra (differential elimination techniques, Gröbner basis methods for noncommutative polynomial rings of ordinary differential operators with polynomial coefficients). The results of [73] have been implemented in the `Maple` package `NonA` built upon the package `OreModules`.

7.2. Research axis 2: Numerical differentiation and finite-time estimation (tools: HOM)

- Algorithms of finite-time and fixed-time observer design have been developed for linear plants based on Implicit Lyapunov function method and homogeneity [30].
- In [23], sufficient conditions for the existence and convergence to zero of numeric approximations to solutions of asymptotically stable homogeneous systems are obtained for the explicit and implicit Euler integration schemes. It is shown that the explicit Euler method has certain drawbacks for the global approximation of homogeneous systems with nonzero degrees, whereas the implicit Euler scheme ensures convergence of the approximating solutions to zero. Properties of absolute and relative errors of the respective discretizations are investigated.
- In [34], the problem of time-varying parameter identification is studied. To this aim, two identification algorithms are developed in order to identify time-varying parameters in a finite-time or prescribed time (fixed-time). The convergence proofs are based on a notion of finite-time stability over finite intervals of time, i.e. Short-finite-time stability; homogeneity for time-varying systems; and Lyapunov-based approach. The results are obtained under injectivity of the regressor term, which is related to the classical identifiability condition. The case of bounded disturbances (noise of measurements) is analyzed for both algorithms. Simulation results illustrate the feasibility of the proposed algorithms.
- [36] contributes to the stability analysis for nonlinear impulsive dynamical systems based on a vector Lyapunov function and its divergence operator. The new method relies on a 2D time domain representation. Different types of stability notions for a class of nonlinear impulsive systems are studied using a vector Lyapunov function approach. The results are applied to analyze the stability of a class of Lipschitz nonlinear impulsive systems based on Linear Matrix Inequalities. Some numerical examples illustrate the feasibility of the proposed approach.
- [21] The rate of convergence to the origin for a chain of integrators stabilized by homogeneous feedback is accelerated by a supervisory switching of control parameters. The proposed acceleration algorithm ensures a fixed-time convergence for otherwise exponentially or finite-time stable homogeneous closed-loop systems. Bounded disturbances are taken into account. The results are especially useful in the case of exponentially stable systems widespread in the practice. The proposed switching strategy is illustrated by computer simulation.

- [37] deals with the design of a robust control for linear systems with external disturbances using a homogeneous differentiator-based observer based on a implicit Lyapunov function approach. Sufficient conditions for stability of the closed-loop system in the presence of external disturbances are obtained and represented by linear matrix inequalities. The parameter tuning for both controller and observer is formulated as a semi-definite programming problem with linear matrix inequalities constraints. Simulation results illustrate the feasibility of the proposed approach and some improvements with respect to the classic linear observer approach.
- Delay estimation algorithms based on sliding mode methodology have been presented in [44].
- A nonlinear distributed observer was proposed in [81] for the problem of distributed estimation in a linear large-scale system.
- In [15], we analyze the observability for linear singular systems with delays, and the corresponding observer design technique has been proposed in [42]. For nonlinear singular system without delay, we propose in [43] a nonlinear Luenberger-like observer. For systems with delays, in [44], we investigate the identifiability of time-delay, and use a sliding mode technique and a classical Newton method to estimate the delay.

7.3. Research axis 3: Control without sophisticated models (tools: ALG-HOM-SET)

- Topological equivalence between quadratically stable and homogeneous asymptotically stable systems have been proven in [72].
- Boundary finite-time control for heat system have been developed in [69], but hyper-exponential control for state delay linear systems have been developed in [70].

7.4. Research axis 4: Applications (tools: ALG-HOM-SET)

- Robust set-point tracking control and optimal control algorithms for turbulent flows have been developed in [26] and tested in Wind Tunnel L1 of ONERA, Lille. (https://www.youtube.com/watch?v=b5NnAV2qeno) The set-point tacking control have been patented, FR 1755440, “Dispositif de contrôle actif du recollement d’un écoulement sur un profil”.
- In [75], [77], the development of a robust (H_∞) control for parametric systems has been initiated. A general framework based on symbolic computation techniques was proposed. In these two papers, the general approach has been applied to the case of linear systems of order up to four and illustrated with the two mass-spring system with damping. In particular, closed forms for the robust controllers and for the robustness radius were obtained. Finally, the robust stabilization of the line of sight of a stabilized mirror system, modeled by a time-delay fourth order system, was studied in [76].
- Within a collaboration with *Safran Tech Laboratory* and *Safran Electronics & Defense*, in [98], we propose a symbolic method for the explicit computation of certain invariant observers studied in navigation theory.
- An experimental synchronization of a family of a recently proposed oscillator model (i.e. the Brockett oscillator) was studied and implemented in [12].
- In [13], high frequency measurements of various water characteristics and nutrients information of the Marel-Carnot sea monitoring station (Boulogne-sur-Mer, France) have been used to identify a physiological model for phytoplankton bloom through the fluorescence signal. An auto-regressive-moving-average with exogenous inputs (ARMAX) model is designed and tested based on the dataset. It was demonstrated that the developed dynamical model can be used for estimating the fluorescence level and for predicting the various states of phytoplankton bloom. Thus, the developed model can be used for monitoring phytoplankton biomass in the water which in turn might give information about unbalanced ecosystem or change in water quality.

- The problem of latency reduction in direct human-computer interaction was considered in [50] and formulated as a trajectory prediction problem. The predictor was constructed as a frequency-domain approximation of the non-casual ideal predictor. This approximation can be computed analytically, or obtained as an optimization task. An adaptive modification of the forecasting algorithm was proposed taking into account possible variations in user behavior.
- In [24], a necessary and sufficient criterion to establish input-to-state stability (ISS) of nonlinear dynamical systems, the dynamics of which are periodic with respect to certain state variables and which possess multiple invariant solutions (equilibria, limit cycles, etc.), is provided. Unlike standard Lyapunov approaches, the condition is relaxed and formulated via a sign-indefinite function with sign-definite derivative, and by taking the system's periodicity explicitly into account. The new result is established by using the framework of cell structure and it complements the ISS theory of multistable dynamics for periodic systems. The efficiency of the proposed approach is illustrated via the global analysis of a nonlinear pendulum with constant persistent input.
- Conditions for almost global stability of an operating point of a realistic model of a synchronous generator with constant field current connected to an infinite bus are derived in [38]. The analysis is conducted by employing the recently proposed concept of input-to-state stability (ISS)–Leonov functions, which is an extension of the powerful cell structure principle developed by Leonov and Noldus to the ISS framework. Compared with the original ideas of Leonov and Noldus, the ISS–Leonov approach has the advantage of providing additional robustness guarantees. The efficiency of the derived sufficient conditions is illustrated via numerical experiments. This article is part of the themed issue 'Energy management: flexibility, risk and optimization'.
- Conditions for existence and global attractivity of the equilibria of a realistic model of a synchronous generator with constant field current connected to an infinite bus are derived in [14]. First, necessary and sufficient conditions for existence and uniqueness of equilibrium points are provided. Then, sufficient conditions for local asymptotic stability and almost global attractivity of one of these equilibria are given. The analysis is carried out by employing a new Lyapunov–like function to establish convergence of bounded trajectories, while the latter is proven using the powerful theoretical framework of cell structures pioneered by Leonov and Noldus. The efficiency of the derived sufficient conditions is illustrated via extensive numerical experiments based on two benchmark examples taken from the literature.
- In [96], we propose a new approach for testing the stability of nD systems. The standard stability conditions are transformed into algebraic conditions and then checked by means of computer algebra techniques for solving algebraic systems such as Gröbner bases, univariate representations and discriminant varieties. The corresponding results were implemented in `Maple`.
- In [17], we address the problem of computing stabilizing controllers for a specific class of multidimensional SISO systems. This problem, which was an open problem (i.e., no effective methods were existing for the computation of stabilizing controllers), has been solved using techniques from computer algebra. As a result, an effective test of stabilizability as well as an algorithm for computing stabilizing controllers were developed.
- We have recently proposed a new method for the *anchor position self-calibration problem*, a rather well-known problem in the signal processing community. In essence, given two sets of wireless communicating devices, i.e. sources and sensors lying in the three dimensional space, the self-calibration algorithm estimates the position of the devices by only using the source–sensor distance measurements. We have first reformulated the problem in terms of certain matrix equalities. They can then be studied in detail by means of computer algebra methods such as Gröbner basis techniques and the package `OreModules`. Coupling symbolic methods with standard linear algebra techniques, we obtain a general solution in all dimensions. In particular, for a space of dimension three, very compact closed-form solutions are obtained in a particular reference frame. Thanks to these closed-form solutions, the noise effect can then be characterized yielding the synthesis of realtime filtering to mitigate the effect of the measurement noise. Finally, the resulting implementation is rather

straightforward and based on real-time operations. Additionally, the underlying numerical tools are standard (least-squares, low-rank factorization, matrix calculus) and well-known. The result of this work is being transferred to a patent. A software prototype `AutoCal` (https://bil.inria.fr/fr/search/query?terms=AutoCal in the BIL) is available on the server `Autocalibrationserver` (https://allgo.inria.fr/webapps/166) under the Inria platform `AllGO`, which allows the user to test the implemented algorithm on his own dataset.

RAPSODI Project-Team

6. New Results

6.1. Variational approach for multiphase flows

In [66], C. Cancès, T. O. Gallouët, and L. Monsaingeon show that the equations governing two-phase flows in porous media have a formal gradient flow structure. The goal of the longer contribution [20] is then twofold. First, it extends the variational interpretation of [66] to the case where an arbitrary number of phases are in competition to flow within a porous medium. Second, we provide rigorous foundations to our claim. More precisely, the convergence of a minimizing movement scheme *à la* Jordan, Kinderlehrer, and Otto [86] is shown in [20], providing by the way a new existence result for multiphase flows in porous media. The result relies on advances tools related to optimal transportation [94], [93].

Based on the previous work, Clément Cancès, Daniel Matthes, and Flore Nabet derived in [46] a model of degenerate Cahn-Hilliard type for the phase segregation in incompressible multiphase flows. The model is obtained as the Wasserstein gradient flow of a Ginzburg-Landau energy with the constraint that the sum of the volume fractions must stay equal to 1. The resulting model differs from the classical degenerate Cahn-Hilliard model (see [97], [77]) and is closely related to a model proposed by Weinan E and collaborators [76], [89]. Besides the derivation of the model, the convergence of a minimizing movement scheme is proven in [46].

6.2. Calculus of variations applied to Image processing, physics and biology

In [23], Benoît Merlet *et al.* consider the branched transportation problem in dimension two with a cost of transport per unit length of path of the form $f_a(m) = a + m$ where $a > 0$ is fixed and m is the flux along the path. As usual in branched transportation, an admissible transport is represented as a vector measure with prescribed divergence $\sum m_j \delta_{x_j} - \sum m'_l \delta_{y_l}$ (the x_j representing the sources and the y_k the sinks). The paper introduces a family of functionals $\{F^a_\varepsilon\}_{\varepsilon>0}$ and the authors establish that this family of functionals approximate the branched transportation energy in the sense of Γ-convergence. The energy F^a_ε is modeled on the Ambrosio-Tortorelli functional and is easy to optimize in practice (using dual formulation for the constraints and alternate direction optimization). In [48], the same authors extend their previous work to functionals defined on k-currents: the objects are no more lines that transport masses but k-dimensional surfaces transporting a given quantity of $(k-1)$-dimensional objects. The ambient space is now of any dimension n. A new family of approximate energies $\{F^a_\varepsilon\}_{\varepsilon>0}$ is introduced and a Γ-convergence analysis is performed in the limit $\varepsilon \downarrow 0$. The limit objects are now k-currents with prescribed boundary, the limit functional controls both their masses (the total flux) and sizes (k-dimensional volume of the object). In the limit $a \downarrow 0$, the limit energy is the k-volume of the object so that these energies can be used for the numerical optimization of the size of k-currents with prescribed boundary. Although rather theoretical, the works [23], [48] are motivated by an image reconstruction issue: how to recover the contours of partially masked objects in an image.

In [26], Michael Goldman and Benoît Merlet study the strong segregation limit for mixtures of Bose-Einstein condensates modelled by a Gross-Pitaievskii functional. They study the behavior of minimizers of the Hamiltonian. First, they show that in the presence of a trapping potential, for different intracomponent strengths, the Thomas-Fermi limit is sufficient to determine the shape of the minimizers. Then they study the case of asymptotically equal intracomponent strengths: at leading order the two phases are then undistinguishable, the authors extract the next order and show that the relevant limit optimization problem is a weighted isoperimetric problem. Then, they study the minimizers, proving radial symmetry or symmetry breaking for different values of the parameters. Eventually, they show that in the absence of a confining potential, even for non-equal intracomponent strengths, one needs to study a related isoperimetric problem to gain information about the shape of the minimizers.

In [49], Michael Goldman, Benoît Merlet and Vincent Millot study a variational problem which models the behavior of topological singularities on the surface of a biological membrane in P_β-phase (see [92]). The problem combines features of the Ginzburg-Landau model in 2D and of the Mumford-Shah functional. As in the classical Ginzburg-Landau theory, a prescribed number of point vortices appear in the moderate energy regime; the model allows for discontinuities, and the energy penalizes their length. The novel phenomenon here is that the vortices have a fractional degree $1/m$ with m prescribed. Those vortices must be connected by line discontinuities to form clusters of total integer degrees. The vortices and line discontinuities are therefore coupled through a topological constraint. As in the Ginzburg-Landau model, the energy is parameterized by a small length scale $\varepsilon > 0$. The authors perform a complete Γ-convergence analysis of the model as $\varepsilon \downarrow 0$ in the moderate energy regime. Then, they study the structure of minimizers of the limit problem. In particular, the line discontinuities of a minimizer solve a variant of the Steiner problem.

6.3. Asymptotic analysis for fluid mechanics

In [28], Ingrid Lacroix-Violet and Alexis Vasseur present the construction of global weak solutions to the quantum Navier-Stokes equation, for any initial value with bounded energy and entropy. The construction is uniform with respect to the Planck constant. This allows to perform the semi-classical limit to the associated compressible Navier-Stokes equation. One of the difficulty of the problem is to deal with the degenerate viscosity, together with the lack of integrability on the velocity. The method is based on the construction of weak solutions that are renormalized in the velocity variable. The existence and stability of these solutions do not need the Mellet-Vasseur inequality.

In [44], the main objective is to generalize to the Navier-Stokes-Korteweg (with density dependent viscosities satisfying the BD relation) and Euler-Korteweg systems a recent relative entropy proposed in [65]. As a concrete application, this helps to justify mathematically the convergence between global weak solutions of the quantum Navier-Stokes system and dissipative solutions of the quantum Euler system when the viscosity coefficient tends to zero. Our results are based on the fact that Euler-Korteweg systems and corresponding Navier–Stokes-Korteweg systems can be reformulated through an augmented system. As a by-product of our analysis, we show that this augmented formulation helps to define relative entropy estimates for the Euler-Korteweg systems in a simplest way and with less hypothesis compared to recent works [74], [80].

In [27], Pierre-Emmanuel Jabin and Thomas Rey investigate the behavior of granular gases in the limit of small Knudsen number, that is, very frequent collisions. They deal with the strongly inelastic case in one dimension of space and velocity. They are able to prove the convergence toward the pressureless Euler system. The proof relies on dispersive relations at the kinetic level, which leads to the so-called Oleinik property at the limit. A more general result is also presented, which can apply to a large class of energy-dissipative kinetic equations.

6.4. Advanced discrete functional analysis results and applications

In [38], Claire Chainais-Hillairet, Benoît Merlet and Alexis Vasseur establish a positive lower bound for the numerical solutions of a stationary convection-diffusion equation on a bounded domain. The proof (which is fully detailed) is based on a celebrated method due to Ennio De Giorgi for showing regularity of the solutions of parabolic and elliptic equations. The robustness of the method allows the authors to adapt it to the discrete solutions obtained by standard finite volume discretizations. Further refinements of this work could lead to improve known error estimates for FV discretizations in L^p-norms to L^∞-norm.

In [14], Marianne Bessemoulin-Chatard and Claire 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 established in [35].

In [43], Marianne Bessemoulin-Chatard and Claire Chainais-Hillairet propose a new proof of existence of a solution to the scheme already introduced in [14] which does not require any assumption on the time step. The result relies on the application of a topological degree argument which is based on the positivity and on uniform-in-time upper bounds of the approximate densities. They also establish uniform-in-time lower bounds satisfied by the approximate densities. These uniform-in-time upper and lower bounds ensure the exponential decay of the scheme towards the thermal equilibrium as shown in [14].

In [12], Boris Andreianov, Clément Cancès, and Ayman Moussa developed a black box to obtain some compactness on the sequence produced by a finite volume discretization for degenerate parabolic problems. Such problems typically appear in the framework of porous media flows or in semi-conductor devices.

6.5. Structure preserving numerical methods

In [7], Clément Cancès and Cindy Guichard proposed in the case of a simple degenerate parabolic equation a nonlinear Control Volume Finite Element (CFVE) scheme that was able to preserve at the discrete level some important features of the continuous problem, namely the positivity of the solution, the decay of the physical energy. The scheme is based on a suitable upwinding procedure and inherits key properties from the Two-Point Flux Approximation (TPFA) finite volume scheme even though the method is not monotone. The convergence of the scheme towards the solution of the continuous problem was also established. In [22], Clément Cancès, Moustafa Ibrahim, and Mazen Saad extend the approach of [7] to the case of the Keller-Segel system with volume filling effect. In [11], Ahmed Ait Hammou Oulhaj, Clément Cancès, and Claire Chainais-Hillairet extend this approach to the Richards equation modeling unsaturated flow in porous media.

In presence of strong anisotropy, the methodology described above may lack robustness: the method is first order accurate, but the error constant may become large in some particularly unfavorable situations. This motivated the development of a new family of schemes with locally positive metric tensor (this denomination was chosen in reference Otto's contribution [90]). The methodology is first developed by Clément Cancès and Cindy Guichard for the so-called Vertex Approximate Gradient (VAG) scheme [79] in [21]. The newly developed method is second order accurate in space and much more robust with respect to the anisotropy than the one of [7] based on upwinding. Then Clément Cancès, Claire Chainais-Hillairet, and Stella Krell extend the methodology to Discrete Duality Finite Volume (DDFV) schemes in [32] and [19].

In [11] (see also the short version [31]), Ahmed Ait Hammou Oulhaj propose an upstream mobility TPFA finite volume scheme for solving a degenerate cross-diffusion problem modeling the flow of two fluids in a porous medium. The scheme has the remarkable property to preserve at the discrete level the local conservation of mass, the positivity of the solution, the decay of the energy. Moreover, the scheme provides a control on the entropy dissipation rate. Thanks to these properties, the convergence of the scheme is established. Numerical simulation show the great robustness of the scheme.

In [37], Clément Cancès and Flore Nabet propose an upstream mobility TPFA finite volume scheme for solving the degenerate Cahn-Hilliard problem. The scheme is designed in order to maintain the positivity of the phase volume fractions, the local conservation of mass and the decay of the energy.

Many applications involve partial differential equations which admits nontrivial steady state solutions. The design of schemes which are able to describe correctly these equilibrium states may be challenging for numerical methods, in particular for high order ones. In [29], inspired by micro-macro decomposition methods for kinetic equations, Lorenzo Pareschi and Thomas Rey present a class of schemes which are capable to preserve the steady state solution and achieve high order accuracy for a class of time dependent partial differential equations including nonlinear diffusion equations and kinetic equations. Extension to systems of conservation laws with source terms are also discussed.

6.6. Numerical approximation of a model for concrete carbonation

In [47], Claire Chainais-Hillairet, Benoît Merlet and Antoine Zurek introduce and study a finite volume scheme for a concrete carbonation model proposed by Aiki and Muntean in [55]. This model consists in a system of two weakly coupled parabolic equations in a varying domain whose length is governed by an ordinary

differential equation. The numerical scheme is obtained by a Euler discretization in time and a Scharfetter-Gummel discretization in space. The convergence of the scheme is established and the existence of a solution to the model is obtained as a by product. Finally, some numerical experiments are performed to show the efficiency of the scheme. The main results of this study are also concisely exposed in [34].

6.7. Numerical methods for stratigraphy problems

In the framework of the PhD thesis of Nicolas Peton, numerical methods are developed for nonlinear diffusion equations arising in stratigraphic modeling. In [33], the special case of a p-Laplacian equation with a constraint on the divergence of the flux is considered. Such a model is used to model erosion and sedimentation processes. The constraint is incorporated to take into account a maximal erosion rate.

6.8. Modeling and numerical simulation of complex fluids

In [25], Giacomo Dimarco, Raphaël Loub ere, Jacek Narski, and Thomas Rey extend the Fast Kinetic Scheme (FKS) originally constructed for solving the BGK equation, to the more challenging case of the Boltzmann equation. The scheme combines a robust and fast method for treating the transport part based on an innovative Lagrangian technique supplemented with conservative fast spectral schemes to treat the collisional operator by means of an operator splitting approach. This approach along with several implementation features related to the parallelization of the algorithm permits to construct an efficient simulation tool which is numerically tested against exact and reference solutions on classical problems arising in rarefied gas dynamic.

In the context of the PhD of Claire Colin-Lecerf, C. Calgaro and co-authors derive in [45] a combined Finite Volumes - Finite Elements (CFVFE) scheme. This work can be seen as a generalization of some previous contributions on incompressible flows [5], [4], [6], in the context of a low-Mach model. Here, the temperature obeying an energy law has been taken into account. The authors chose to solve the continuity equation and the state equation linking temperature, density and thermodynamic pressure is imposed implicitly. Now the velocity field is no more divergence-free, so that the projection method solving the momentum equation has to be adapted. This combined scheme preserve the constant state and ensure the discrete maximum principle on the density. Their numerical results have been compared to some others which use purely finite elements schemes (see [62], [58], [81]) and in particular on a benchmark consisting in a transient hot jet entering in a cavity.

Diffuse interface models, such as the Kazhikhov-Smagulov model, allow to describe some phase transition phenomena. The theoretical analysis of this model was given by Bresch at al. [64] (see also reference therein). In the previous work [6], C. Calgaro et al. have implemented the CFVFE scheme and studied numerically the progression of the front of a powder-snow avalanche with respect to some characteristics parameters of the flow, such as the Froude, Schmidt and Reynolds numbers. In [18], C. Calgaro and co-authors investigate theoretically the CFVFE scheme. They construct a fully discrete numerical scheme for approximating the two-dimensional Kazhikhov-Smagulov model, using a first-order time discretization and a splitting in time to allow the construction of the combined scheme. Consequently, at each time step, one only needs to solve two decoupled problems, the first one for the density (using the Finite Volume method) and the second one for the velocity and pressure (using the Finite Element method). The authors prove the stability of the combined scheme and the convergence towards the global in time weak solution of the model. In this model, the convection-diffusion equation for the density can also be discretized by a implicit-explicit (IMEX) second order method in the Finite Volume scheme. In the framework of MUSCL methods, C. Calgaro and M. Ezzoug prove in [36] that the local maximum property is guaranteed under an explicit Courant-Friedrichs-Levy condition and the classical hypothesis for the triangulation of the domain.

6.9. Cost reduction of numerical methods

This section gathers contributions for which the main motivation was to increase the efficiency of numerical methods, either by localizing the computational effort thanks to mesh refinement

In [24], E. Creusé and his collaborators generalize the equilibrated error estimators developed in the low-frequency magnetostatic case to the case of the harmonic time-dependent one. This contribution allows to obtain a bound of the numerical error equal to one, so that the accuracy of the obtained solution can be explicitly controlled.

The contribution [16] by K. Brenner and C. Cancès is devoted to the improvement of the behavior of Newton's method when solving degenerate parabolic equations. Such equations are very common for instance in the context of complex porous media flows. In [16], the presentation focuses on Richards equation modeling saturated/unsaturated flows in porous media. The basic idea is the following: Newton's method is not invariant by nonlinear change of variables. The choice of the primary variable then impacts the effective resolution of the nonlinear system provided by the scheme. The idea developed in [16] is then to construct an abstract primary variable to facilitate Newton's method's convergence. This leads to an impressive reduction of the computational cost, a better accuracy in the results and a strong robustness of the method w.r.t. the nonlinearities appearing in the continuous model.

In [39], Ward Melis, Thomas Rey, and Giovanni Samaey present a high-order, fully explicit, asymptotic-preserving projective integration scheme for the nonlinear BGK equation. The method first takes a few small (inner) steps with a simple, explicit method (such as direct forward Euler) to damp out the stiff components of the solution. Then, the time derivative is estimated and used in an (outer) Runge-Kutta method of arbitrary order. Based on the spectrum of the linearized BGK operator, they deduce that, with an appropriate choice of inner step size, the time step restriction on the outer time step as well as the number of inner time steps is independent of the stiffness of the BGK source term. They illustrate the method with numerical results in one and two spatial dimensions.

In [13], Christophe Besse, Guillaume Dujardin, and Ingrid Lacroix-Violet present the numerical integration in time of nonlinear Schrödinger equations with rotating term. After performing a change of unknown so that the rotation term disappears they consider exponential integrators such as exponential Runge-Kutta methods and Lawson methods. They provide an analysis of the order of convergence and some preservation properties of these methods and they present numerical experiments.

SEQUEL Project-Team

7. New Results

7.1. Decision-making Under Uncertainty

7.1.1. Reinforcement Learning

Thompson Sampling for Linear-Quadratic Control Problems, [22]

We consider the exploration-exploitation tradeoff in linear quadratic (LQ) control problems, where the state dynamics is linear and the cost function is quadratic in states and controls. We analyze the regret of Thompson sampling (TS) (a.k.a. posterior-sampling for reinforcement learning) in the frequentist setting, i.e., when the parameters characterizing the LQ dynamics are fixed. Despite the empirical and theoretical success in a wide range of problems from multi-armed bandit to linear bandit, we show that when studying the frequentist regret TS in control problems, we need to trade-off the frequency of sampling optimistic parameters and the frequency of switches in the control policy. This results in an overall regret of $O(T^{2/3})$, which is significantly worse than the regret $O(\sqrt{T})$ achieved by the optimism-in-face-of-uncertainty algorithm in LQ control problems.

Exploration–Exploitation in MDPs with Options, [33]

While a large body of empirical results show that temporally-extended actions and options may significantly affect the learning performance of an agent, the theoretical understanding of how and when options can be beneficial in online reinforcement learning is relatively limited. In this paper, we derive an upper and lower bound on the regret of a variant of UCRL using options. While we first analyze the algorithm in the general case of semi-Markov decision processes (SMDPs), we show how these results can be translated to the specific case of MDPs with options and we illustrate simple scenarios in which the regret of learning with options can be provably much smaller than the regret suffered when learning with primitive actions.

Regret Minimization in MDPs with Options without Prior Knowledge, [34]

The option framework integrates temporal abstraction into the reinforcement learning model through the introduction of macro-actions (i.e., options). Recent works leveraged the mapping of Markov decision processes (MDPs) with options to semi-MDPs (SMDPs) and introduced SMDP-versions of exploration-exploitation algorithms (e.g., RMAX-SMDP and UCRL-SMDP) to analyze the impact of options on the learning performance. Nonetheless, the PAC-SMDP sample complexity of RMAX-SMDP can hardly be translated into equivalent PAC-MDP theoretical guarantees, while the regret analysis of UCRL-SMDP requires prior knowledge of the distributions of the cumulative reward and duration of each option, which are hardly available in practice. In this paper, we remove this limitation by combining the SMDP view together with the inner Markov structure of options into a novel algorithm whose regret performance matches UCRL-SMDP's up to an additive regret term. We show scenarios where this term is negligible and the advantage of temporal abstraction is preserved. We also report preliminary empirical results supporting the theoretical findings.

Is the Bellman Residual a Bad Proxy?, [36]

This paper aims at theoretically and empirically comparing two standard optimization criteria for Reinforcement Learning: i) maximization of the mean value and ii) minimization of the Bellman residual. For that purpose, we place ourselves in the framework of policy search algorithms, that are usually designed to maximize the mean value, and derive a method that minimizes the residual T * v π – v π 1,ν over policies. A theoretical analysis shows how good this proxy is to policy optimization , and notably that it is better than its value-based counterpart. We also propose experiments on randomly generated generic Markov decision processes, specifically designed for studying the influence of the involved concentrability coefficient. They show that the Bellman residual is generally a bad proxy to policy optimization and that directly maximizing the mean value is much better, despite the current lack of deep theoretical analysis. This might seem obvious,

as directly addressing the problem of interest is usually better, but given the prevalence of (projected) Bellman residual minimization in value-based reinforcement learning, we believe that this question is worth to be considered.

Faut-il minimiser le résidu de Bellman ou maximiser la valeur moyenne ?, [56]

Transfer Reinforcement Learning with Shared Dynamics, [38]

This article addresses a particular Transfer Reinforcement Learning (RL) problem: when dynamics do not change from one task to another, and only the reward function does. Our method relies on two ideas, the first one is that transition samples obtained from a task can be reused to learn on any other task: an immediate reward estimator is learnt in a supervised fashion and for each sample, the reward entry is changed by its reward estimate. The second idea consists in adopting the optimism in the face of uncertainty principle and to use upper bound reward estimates. Our method is tested on a navigation task, under four Transfer RL experimental settings: with a known reward function, with strong and weak expert knowledge on the reward function, and with a completely unknown reward function. It is also evaluated in a Multi-Task RL experiment and compared with the state-of-the-art algorithms. Results reveal that this method constitutes a major improvement for transfer/multi-task problems that share dynamics.

7.1.2. Multi-arm Bandit Theory

Trading Off Rewards and Errors in Multi-armed Bandits, [31]

In multi-armed bandits, the most common objective is the maximization of the cumulative reward. Alternative settings include active exploration, where a learner tries to gain accurate estimates of the rewards of all arms. While these objectives are contrasting, in many scenarios it is desirable to trade off rewards and errors. For instance, in educational games the designer wants to gather generalizable knowledge about the behavior of the students and teaching strategies (small estimation errors) but, at the same time, the system needs to avoid giving a bad experience to the players, who may leave the system permanently (large reward). In this paper, we formalize this tradeoff and introduce the ForcingBalance algorithm whose performance is provably close to the best possible tradeoff strategy. Finally, we demonstrate on real-world educational data that ForcingBalance returns useful information about the arms without compromising the overall reward.

Online Influence Maximization Under Independent Cascade Model with Semi-bandit Feedback, [54]

We study the online influence maximization problem in social networks under the independent cascade model. Specifically, we aim to learn the set of " best influencers " in a social network online while repeatedly interacting with it. We address the challenges of (i) combinatorial action space, since the number of feasible influencer sets grows exponentially with the maximum number of influencers, and (ii) limited feedback, since only the influenced portion of the network is observed. Under a stochastic semi-bandit feedback, we propose and analyze IMLinUCB, a computationally efficient UCB-based algorithm. Our bounds on the cumulative regret are polynomial in all quantities of interest, achieve near-optimal dependence on the number of interactions and reflect the topology of the network and the activation probabilities of its edges, thereby giving insights on the problem complexity. To the best of our knowledge, these are the first such results. Our experiments show that in several representative graph topologies, the regret of IMLinUCB scales as suggested by our upper bounds. IMLinUCB permits linear generalization and thus is both statistically and computationally suitable for large-scale problems. Our experiments also show that IMLinUCB with linear generalization can lead to low regret in real-world online influence maximization.

Boundary Crossing for General Exponential Families, [39]

We consider parametric exponential families of dimension K on the real line. We study a variant of boundary crossing probabilities coming from the multi-armed bandit literature, in the case when the real-valued distributions form an exponential family of dimension K. Formally, our result is a concentration inequality that bounds the probability that B ψ (θ n , θ) f (t/n)/n, where θ is the parameter of an unknown target distribution, θ n is the empirical parameter estimate built from n observations, ψ is the log-partition function of the exponential family and B ψ is the corresponding Bregman divergence. From the perspective of stochastic multi-armed bandits, we pay special attention to the case when the boundary function f is logarithmic, as it enables to

analyze the regret of the state-of-the-art KL-ucb and KL-ucb+ strategies, whose analysis was left open in such generality. Indeed, previous results only hold for the case when K = 1, while we provide results for arbitrary finite dimension K, thus considerably extending the existing results. Perhaps surprisingly, we highlight that the proof techniques to achieve these strong results already existed three decades ago in the work of T.L. Lai, and were apparently forgotten in the bandit community. We provide a modern rewriting of these beautiful techniques that we believe are useful beyond the application to stochastic multi-armed bandits.

The Non-stationary Stochastic Multi-armed Bandit Problem, Robin, Féraud, Maillard [64] [0]

Linear Thompson Sampling Revisited, [21]

We derive an alternative proof for the regret of Thompson sampling (TS) in the stochastic linear bandit setting. While we obtain a regret bound of order $\widetilde{O}(d^{3/2}\sqrt{T})$ as in previous results, the proof sheds new light on the functioning of the TS. We leverage on the structure of the problem to show how the regret is related to the sensitivity (i.e., the gradient) of the objective function and how selecting optimal arms associated to *optimistic* parameters does control it. Thus we show that TS can be seen as a generic randomized algorithm where the sampling distribution is designed to have a fixed probability of being optimistic, at the cost of an additional $\sqrt{d}$ regret factor compared to a UCB-like approach. Furthermore, we show that our proof can be readily applied to regularized linear optimization and generalized linear model problems.

Active Learning for Accurate Estimation of Linear Models, [47]

We explore the sequential decision-making problem where the goal is to estimate a number of linear models uniformly well, given a shared budget of random contexts independently sampled from a known distribution. For each incoming context, the decision-maker selects one of the linear models and receives an observation that is corrupted by the unknown noise level of that model. We present Trace-UCB, an adaptive allocation algorithm that learns the models' noise levels while balancing contexts accordingly across them, and prove bounds for its simple regret in both expectation and high-probability. We extend the algorithm and its bounds to the high dimensional setting , where the number of linear models times the dimension of the contexts is more than the total budget of samples. Simulations with real data suggest that Trace-UCB is remarkably robust , outperforming a number of baselines even when its assumptions are violated.

Learning the Distribution with Largest Mean: Two Bandit Frameworks, [18]

Over the past few years, the multi-armed bandit model has become increasingly popular in the machine learning community, partly because of applications including online content optimization. This paper reviews two different sequential learning tasks that have been considered in the bandit literature ; they can be formulated as (sequentially) learning which distribution has the highest mean among a set of distributions, with some constraints on the learning process. For both of them (regret minimization and best arm identification) we present recent, asymptotically optimal algorithms. We compare the behaviors of the sampling rule of each algorithm as well as the complexity terms associated to each problem.

On Bayesian Index Policies for Sequential Resource Allocation, [19]

This paper is about index policies for minimizing (frequentist) regret in a stochastic multi-armed bandit model, inspired by a Bayesian view on the problem. Our main contribution is to prove that the Bayes-UCB algorithm, which relies on quantiles of posterior distributions, is asymptotically optimal when the reward distributions belong to a one-dimensional exponential family, for a large class of prior distributions. We also show that the Bayesian literature gives new insight on what kind of exploration rates could be used in frequentist, UCB-type algorithms. Indeed, approximations of the Bayesian optimal solution or the Finite Horizon Gittins indices provide a justification for the kl-UCB+ and kl-UCB-H+ algorithms, whose asymptotic optimality is also established.

Multi-Player Bandits Models Revisited, [59]

Multi-player Multi-Armed Bandits (MAB) have been extensively studied in the literature, motivated by applications to Cognitive Radio systems. Driven by such applications as well, we motivate the introduction

[0] This work has been done while OA. Maillard was at Inria Saclay, in the TAO team.

of several levels of feedback for multi-player MAB algorithms. Most existing work assume that sensing information is available to the algorithm. Under this assumption, we improve the state-of-the-art lower bound for the regret of any decentralized algorithms and introduce two algorithms, RandTopM and MCTopM, that are shown to empirically outperform existing algorithms. Moreover, we provide strong theoretical guarantees for these algorithms, including a notion of asymptotic optimality in terms of the number of selections of bad arms. We then introduce a promising heuristic, called Selfish, that can operate without sensing information, which is crucial for emerging applications to Internet of Things networks. We investigate the empirical performance of this algorithm and provide some first theoretical elements for the understanding of its behavior.

Multi-Armed Bandit Learning in IoT Networks: Learning helps even in non-stationary settings, [57]

Setting up the future Internet of Things (IoT) networks will require to support more and more communicating devices. We prove that intelligent devices in unlicensed bands can use Multi-Armed Bandit (MAB) learning algorithms to improve resource exploitation. We evaluate the performance of two classical MAB learning algorithms, UCB1 and Thompson Sampling, to handle the decentralized decision-making of Spectrum Access, applied to IoT networks; as well as learning performance with a growing number of intelligent end-devices. We show that using learning algorithms does help to fit more devices in such networks, even when all end-devices are intelligent and are dynamically changing channel. In the studied scenario, stochastic MAB learning provides a up to 16% gain in term of successful transmission probabilities, and has near optimal performance even in non-stationary and non-i.i.d. settings with a majority of intelligent devices.

7.1.3. Nonparametric Statistics of Time Series

Efficient Tracking of a Growing Number of Experts, [41]

We consider a variation on the problem of prediction with expert advice, where new forecasters that were unknown until then may appear at each round. As often in prediction with expert advice, designing an algorithm that achieves near-optimal regret guarantees is straightforward, using aggregation of experts. However, when the comparison class is sufficiently rich, for instance when the best expert and the set of experts itself changes over time, such strategies naively require to maintain a prohibitive number of weights (typically exponential with the time horizon). By contrast, designing strategies that both achieve a near-optimal regret and maintain a reasonable number of weights is highly non-trivial. We consider three increasingly challenging objectives (simple regret, shifting regret and sparse shifting regret) that extend existing notions defined for a fixed expert ensemble; in each case, we design strategies that achieve tight regret bounds, adaptive to the parameters of the comparison class, while being computationally inexpensive. Moreover, our algorithms are anytime , agnostic to the number of incoming experts and completely parameter-free. Such remarkable results are made possible thanks to two simple but highly effective recipes: first the " abstention trick " that comes from the specialist framework and enables to handle the least challenging notions of regret, but is limited when addressing more sophisticated objectives. Second, the " muting trick " that we introduce to give more flexibility. We show how to combine these two tricks in order to handle the most challenging class of comparison strategies.

7.1.4. Stochastic Games

Monte-Carlo Tree Search by Best Arm Identification, [37]

Recent advances in bandit tools and techniques for sequential learning are steadily enabling new applications and are promising the resolution of a range of challenging related problems. We study the game tree search problem, where the goal is to quickly identify the optimal move in a given game tree by sequentially sampling its stochastic payoffs. We develop new algorithms for trees of arbitrary depth, that operate by summarizing all deeper levels of the tree into confidence intervals at depth one, and applying a best arm identification procedure at the root. We prove new sample complexity guarantees with a refined dependence on the problem instance. We show experimentally that our algorithms outperform existing elimination-based algorithms and match previous special-purpose methods for depth-two trees.

Learning Nash Equilibrium for General-Sum Markov Games from Batch Data, [46]

This paper addresses the problem of learning a Nash equilibrium in γ-discounted mul-tiplayer general-sum Markov Games (MGs) in a batch setting. As the number of players increases in MG, the agents may either collaborate or team apart to increase their final rewards. One solution to address this problem is to look for a Nash equilibrium. Although , several techniques were found for the subcase of two-player zero-sum MGs, those techniques fail to find a Nash equilibrium in general-sum Markov Games. In this paper, we introduce a new definition of-Nash equilibrium in MGs which grasps the strategy's quality for multiplayer games. We prove that minimizing the norm of two Bellman-like residuals implies to learn such an-Nash equilibrium. Then, we show that minimizing an empirical estimate of the L p norm of these Bellman-like residuals allows learning for general-sum games within the batch setting. Finally, we introduce a neural network architecture that successfully learns a Nash equilibrium in generic multiplayer general-sum turn-based MGs.

7.1.5. Automata Learning

Spectral Learning from a Single Trajectory under Finite-State Policies, [23]

We present spectral methods of moments for learning sequential models from a single trajectory, in stark contrast with the classical literature that assumes the availability of multiple i.i.d. trajectories. Our approach leverages an efficient SVD-based learning algorithm for weighted automata and provides the first rigorous analysis for learning many important models using dependent data. We state and analyze the algorithm under three increasingly difficult scenarios: probabilistic automata, stochastic weighted automata, and reactive predictive state representations controlled by a finite-state policy. Our proofs include novel tools for studying mixing properties of stochastic weighted automata.

7.1.6. Online Kernel and Graph-Based Methods

Distributed Adaptive Sampling for Kernel Matrix Approximation, [26]

Most kernel-based methods, such as kernel regression, kernel PCA, ICA, or k-means clustering, do not scale to large datasets, because constructing and storing the kernel matrix $\mathbf{K}_n$ requires at least $O(n^2)$ time and space for n samples. Recent works (Alaoui 2014, Musco 2016) show that sampling points with replacement according to their ridge leverage scores (RLS) generates small dictionaries of relevant points with strong spectral approximation guarantees for $\mathbf{K}_n$. The drawback of RLS-based methods is that computing exact RLS requires constructing and storing the whole kernel matrix. In this paper, we introduce SQUEAK, a new algorithm for kernel approximation based on RLS sampling that sequentially processes the dataset, storing a dictionary which creates accurate kernel matrix approximations with a number of points that only depends on the effective dimension $d_{\text{eff}}(\gamma)$ of the dataset. Moreover since all the RLS estimations are efficiently performed using only the small dictionary, SQUEAK never constructs the whole matrix $\mathbf{K}_n$ runs in linear time $\widetilde{O}(nd_{\text{eff}}(\gamma)^3)$ w.r.t. n, and requires only a single pass over the dataset. We also propose a parallel and distributed version of SQUEAK achieving similar accuracy in as little as $\widetilde{O}(log(n)d_{\text{eff}}(\gamma)^3)$ time.

Second-Order Kernel Online Convex Optimization with Adaptive Sketching, [28]

Kernel online convex optimization (KOCO) is a framework combining the expressiveness of non-parametric kernel models with the regret guarantees of online learning. First-order KOCO methods such as functional gradient descent require only $O(t)$ time and space per iteration, and, when the only information on the losses is their convexity, achieve a minimax optimal $O(\sqrt{T})$ regret. Nonetheless, many common losses in kernel problems, such as squared loss, logistic loss, and squared hinge loss posses stronger curvature that can be exploited. In this case, second-order KOCO methods achieve $O(\log(Det(K)))$ regret, which we show scales as $O(d_{eff}\log T)$, where d_{eff} is the effective dimension of the problem and is usually much smaller than $O(\sqrt{T})$. The main drawback of second-order methods is their much higher $O(t^2)$ space and time complexity. In this paper, we introduce kernel online Newton step (KONS), a new second-order KOCO method that also achieves $O(d_{eff}\log T)$ regret. To address the computational complexity of second-order methods, we introduce a new matrix sketching algorithm for the kernel matrix K, and show that for a chosen parameter $\gamma \leq 1$ our Sketched-KONS reduces the space and time complexity by a factor of γ^2 to $O(t^2\gamma^2)$ space and time per iteration, while incurring only $1/\gamma$ times more regret.

Efficient Second-order Online Kernel Learning with Adaptive Embedding, [27]

Online kernel learning (OKL) is a flexible framework to approach prediction problems, since the large approximation space provided by reproducing kernel Hilbert spaces can contain an accurate function for the problem. Nonetheless, optimizing over this space is computationally expensive. Not only first order methods accumulate $O(\sqrt{T})$ more loss than the optimal function, but the curse of kernelization results in a $O(t)$ per step complexity. Second-order methods get closer to the optimum much faster, suffering only $O(\log(T))$ regret, but second-order updates are even more expensive, with a $O(t^2)$ per-step cost. Existing approximate OKL methods try to reduce this complexity either by limiting the Support Vectors (SV) introduced in the predictor, or by avoiding the kernelization process altogether using embedding. Nonetheless, as long as the size of the approximation space or the number of SV does not grow over time, an adversary can always exploit the approximation process. In this paper, we propose PROS-N-KONS, a method that combines Nystrom sketching to project the input point in a small, accurate embedded space, and performs efficient second-order updates in this space. The embedded space is continuously updated to guarantee that the embedding remains accurate, and we show that the per-step cost only grows with the effective dimension of the problem and not with T. Moreover, the second-order updated allows us to achieve the logarithmic regret. We empirically compare our algorithm on recent large-scales benchmarks and show it performs favorably.

Zonotope Hit-and-run for Efficient Sampling from Projection DPPs, [35]

Determinantal point processes (DPPs) are distributions over sets of items that model diversity using kernels. Their applications in machine learning include summary extraction and recommendation systems. Yet, the cost of sampling from a DPP is prohibitive in large-scale applications, which has triggered an effort towards efficient approximate samplers. We build a novel MCMC sampler that combines ideas from combinatorial geometry, linear programming, and Monte Carlo methods to sample from DPPs with a fixed sample cardinality, also called projection DPPs. Our sampler leverages the ability of the hit-and-run MCMC kernel to efficiently move across convex bodies. Previous theoretical results yield a fast mixing time of our chain when targeting a distribution that is close to a projection DPP, but not a DPP in general. Our empirical results demonstrate that this extends to sampling projection DPPs, i.e., our sampler is more sample-efficient than previous approaches which in turn translates to faster convergence when dealing with costly-to-evaluate functions, such as summary extraction in our experiments.

7.2. Statistical Learning and Bayesian Analysis

Universality of Bayesian mixture predictors, [50]

The problem is that of sequential probability forecasting for finite-valued time series. The data is generated by an unknown probability distribution over the space of all one-way infinite sequences. It is known that this measure belongs to a given set C, but the latter is completely arbitrary (uncountably infinite, without any structure given). The performance is measured with asymptotic average log loss. In this work it is shown that the minimax asymptotic performance is always attainable, and it is attained by a convex combination of a countably many measures from the set C (a Bayesian mixture). This was previously only known for the case when the best achievable asymptotic error is 0. This also contrasts previous results that show that in the non-realizable case all Bayesian mixtures may be suboptimal, while there is a predictor that achieves the optimal performance.

Hypotheses Testing on Infinite Random Graphs, [48]

Drawing on some recent results that provide the formalism necessary to definite stationarity for infinite random graphs, this paper initiates the study of statistical and learning questions pertaining to these objects. Specifically, a criterion for the existence of a consistent test for complex hypotheses is presented, generalizing the corresponding results on time series. As an application, it is shown how one can test that a tree has the Markov property, or,more generally, to estimate its memory.

Independence Clustering (Without a Matrix), [49]

The independence clustering problem is considered in the following formulation: given a set S of random variables, it is required to find the finest partitioning $\{U\_1, \cdots, U\_k\}$ of S into clusters such that the clusters $U\_1, \cdots, U\_k$ are mutually independent. Since mutual independence is the target, pairwise similarity

measurements are of no use, and thus traditional clustering algorithms are inapplicable. The distribution of the random variables in S is, in general, unknown, but a sample is available. Thus, the problem is cast in terms of time series. Two forms of sampling are considered: i.i.d. and stationary time series, with the main emphasis being on the latter, more general, case. A consistent, computationally tractable algorithm for each of the settings is proposed, and a number of open directions for further research are outlined.

7.3. Applications

7.3.1. Dialogue Systems and Natural Language

End-to-end Optimization of Goal-driven and Visually Grounded Dialogue Systems, [51]

End-to-end design of dialogue systems has recently become a popular research topic thanks to powerful tools such as encoder-decoder architectures for sequence-to-sequence learning. Yet, most current approaches cast human-machine dialogue management as a supervised learning problem, aiming at predicting the next utterance of a participant given the full history of the dialogue. This vision is too simplistic to render the intrinsic planning problem inherent to dialogue as well as its grounded nature , making the context of a dialogue larger than the sole history. This is why only chitchat and question answering tasks have been addressed so far using end-to-end architectures. In this paper, we introduce a Deep Reinforcement Learning method to optimize visually grounded task-oriented dialogues , based on the policy gradient algorithm. This approach is tested on a dataset of 120k dialogues collected through Mechanical Turk and provides encouraging results at solving both the problem of generating natural dialogues and the task of discovering a specific object in a complex picture.

Online Learning and Transfer for User Adaptation in Dialogue Systems, [58]

We address the problem of user adaptation in Spoken Dialogue Systems. The goal is to quickly adapt online to a new user given a large amount of dialogues collected with other users. Previous works using Transfer for Reinforcement Learning tackled this problem when the number of source users remains limited. In this paper, we overcome this constraint by clustering the source users: each user cluster, represented by its centroid, is used as a potential source in the state-of-the-art Transfer Reinforcement Learning algorithm. Our benchmark compares several clustering approaches , including one based on a novel metric. All experiments are led on a negotiation dialogue task, and their results show significant improvements over baselines.

GuessWhat?! Visual Object Discovery Through Multi-modal Dialogue, [29]

We introduce GuessWhat?!, a two-player guessing game as a testbed for research on the interplay of computer vision and dialogue systems. The goal of the game is to locate an unknown object in a rich image scene by asking a sequence of questions. Higher-level image understanding, like spatial reasoning and language grounding, is required to solve the proposed task. Our key contribution is the collection of a large-scale dataset consisting of 150K human-played games with a total of 800K visual question-answer pairs on 66K images. We explain our design decisions in collecting the dataset and introduce the oracle and questioner tasks that are associated with the two players of the game. We prototyped deep learning models to establish initial base-lines of the introduced tasks.

LIG-CRIStAL System for the WMT17 Automatic Post-Editing Task, [25]

This paper presents the LIG-CRIStAL submission to the shared Automatic Post-Editing task of WMT 2017. We propose two neural post-editing models: a mono-source model with a task-specific attention mechanism, which performs particularly well in a low-resource scenario; and a chained architecture which makes use of the source sentence to provide extra context. This latter architecture manages to slightly improve our results when more training data is available. We present and discuss our results on two datasets (en-de and de-en) that are made available for the task.

7.3.2. Recommendation systems

A Multi-Armed Bandit Model Selection for Cold-Start User Recommendation, [32]

How can we effectively recommend items to a user about whom we have no information? This is the problem we focus on, known as the cold-start problem. In this paper, we focus on the cold user problem.In most existing works, the cold-start problem is handled through the use of many kinds of information available about the user. However, what happens if we do not have any information?Recommender systems usually keep a substantial amount of prediction models that are available for analysis. Moreover, recommendations to new users yield uncertain returns. Assuming a number of alternative prediction models is available to select items to recommend to a cold user, this paper introduces a multi-armed bandit based model selection, named PdMS.In comparison with two baselines, PdMS improves the performance as measured by the nDCG.These improvements are demonstrated on real, public datasets.

7.3.3. Software development

A Large-scale Study of Call Graph-based Impact Prediction using Mutation Testing, [20]

In software engineering, impact analysis consists in predicting the software elements (e.g. modules, classes, methods) potentially impacted by a change in the source code. Impact analysis is required to optimize the testing effort. In this paper, we propose a framework to predict error propagation. Based on 10 open-source Java projects and 5 classical mutation operators, we create 17000 mutants and study how the error they introduce propagates. This framework enables us to analyze impact prediction based on four types of call graph. Our results show that the sophistication indeed increases completeness of impact prediction. However, and surprisingly to us, the most basic call graph gives the highest trade-off between precision and recall for impact prediction.

Correctness Attraction: A Study of Stability of Software Behavior under Runtime Perturbation, [15]

Can the execution of a software be perturbed without breaking the correctness of the output? In this paper, we devise a novel protocol to answer this rarely investigated question. In an experimental study, we observe that many perturbations do not break the correctness in ten subject programs. We call this phenomenon “correctness attraction”. The uniqueness of this protocol is that it considers a systematic exploration of the perturbation space as well as perfect oracles to determine the correctness of the output. To this extent, our findings on the stability of software under execution perturbations have a level of validity that has never been reported before in the scarce related work. A qualitative manual analysis enables us to set up the first taxonomy ever of the reasons behind correctness attraction.

7.3.4. Graph theory

A generative model for sparse, evolving digraphs, [43]

Generating graphs that are similar to real ones is an open problem, while the similarity notion is quite elusive and hard to formalize. In this paper, we focus on sparse digraphs and propose SDG, an algorithm that aims at generating graphs similar to real ones. Since real graphs are evolving and this evolution is important to study in order to understand the underlying dynamical system, we tackle the problem of generating series of graphs. We propose SEDGE, an algorithm meant to generate series of graphs similar to a real series. SEDGE is an extension of SDG. We consider graphs that are representations of software programs and show experimentally that our approach outperforms other existing approaches. Experiments show the performance of both algorithms.

A Spectral Algorithm with Additive Clustering for the Recovery of Overlapping Communities in Networks, [17]

This paper presents a novel spectral algorithm with additive clustering designed to identify overlapping communities in networks. The algorithm is based on geometric properties of the spectrum of the expected adjacency matrix in a random graph model that we call stochastic blockmodel with overlap (SBMO). An adaptive version of the algorithm, that does not require the knowledge of the number of hidden communities, is proved to be consistent under the SBMO when the degrees in the graph are (slightly more than) logarithmic. The algorithm is shown to perform well on simulated data and on real-world graphs with known overlapping communities.

7.3.5. Deep Learning

Modulating early visual processing by language, [30]

It is commonly assumed that language refers to high-level visual concepts while leaving low-level visual processing unaffected. This view dominates the current literature in computational models for language-vision tasks, where visual and linguistic inputs are mostly processed independently before being fused into a single representation. In this paper, we deviate from this classic pipeline and propose to modulate the entire visual processing by a linguistic input. Specifically, we introduce Conditional Batch Normalization (CBN) as an efficient mechanism to modulate convolutional feature maps by a linguistic embedding. We apply CBN to a pre-trained Residual Network (ResNet), leading to the MODulatEd ResNet (MODERN) architecture, and show that this significantly improves strong baselines on two visual question answering tasks. Our ablation study confirms that modulating from the early stages of the visual processing is beneficial.

FiLM: Visual Reasoning with a General Conditioning Layer, [45]

We introduce a general-purpose conditioning method for neural networks called FiLM: Feature-wise Linear Modulation. FiLM layers influence neural network computation via a simple , feature-wise affine transformation based on conditioning information. We show that FiLM layers are highly effective for visual reasoning - answering image-related questions which require a multi-step, high-level process - a task which has proven difficult for standard deep learning methods that do not explicitly model reasoning. Specifically, we show on visual reasoning tasks that FiLM layers 1) halve state-of-the-art error for the CLEVR benchmark, 2) modulate features in a coherent manner, 3) are robust to ablations and architectural modifications, and 4) generalize well to challenging, new data from few examples or even zero-shot.

Learning Visual Reasoning Without Strong Priors, [44]

Achieving artificial visual reasoning - the ability to answer image-related questions which require a multi-step, high-level process - is an important step towards artificial general intelligence. This multi-modal task requires learning a question-dependent, structured reasoning process over images from language. Standard deep learning approaches tend to exploit biases in the data rather than learn this underlying structure, while leading methods learn to visually reason successfully but are hand-crafted for reasoning. We show that a general-purpose, Conditional Batch Normalization approach achieves state-of-the-art results on the CLEVR Visual Reasoning benchmark with a 2.4% error rate. We outperform the next best end-to-end method (4.5%) and even methods that use extra supervision (3.1%). We probe our model to shed light on how it reasons, showing it has learned a question-dependent, multi-step process. Previous work has operated under the assumption that visual reasoning calls for a specialized architecture, but we show that a general architecture with proper conditioning can learn to visually reason effectively. Index Terms: Deep Learning, Language and Vision Note: A full paper extending this study is available at http: //arxiv.org/abs/1709.07871, with additional references , experiments, and analysis.

HoME: a Household Multimodal Environment, [24]

We introduce HoME: a Household Multimodal Environment for artificial agents to learn from vision, audio, semantics, physics, and interaction with objects and other agents, all within a realistic context. HoME integrates over 45,000 diverse 3D house layouts based on the SUNCG dataset, a scale which may facilitate learning, generalization, and transfer. HoME is an open-source, OpenAI Gym-compatible platform extensible to tasks in reinforcement learning, language grounding, sound-based navigation, robotics, multi-agent learning, and more. We hope HoME better enables artificial agents to learn as humans do: in an interactive, multimodal, and richly contextualized setting.

BONSAI Project-Team

7. New Results

7.1. Metagenomics

Reconstruction of phylogenetic marker genes. Accurate identification of organisms present within a community is essential to understanding the structure of an ecosystem. However, current HTS technologies generate short reads, such as Illumina reads, which makes it a difficult task. One possibility is to focus on assembly of taxonomic markers of interest, such as 16S ribosomal RNA. The PhD thesis of P. Pericard proposed an algorithm that is specifically dedicated to this problem. The method implements a stepwise process based on construction and analysis of a read overlap graph, which is built using read alignments (produced by SortMeRNA) and is decomposed into relevant connected components extracted from a compressed representation of the graph. It is able to recover full length 16S sequences with high precision assemblies ($\leq 0.1\%$ error rate). This work is published in the reference journal in the field [23] and the resulting software, MATAM, was released this spring. It is currently being tested in several labs [0]. This work received the Best Oral Presentation Award from the SFBI [0] this year [29].

Metagenomics assembly. Another important task that could help taxonomic assignment is to reconstruct uncultured microbial strains and species for which the genome sequence is fully unknown. To this end, metagenomics mainly borrows techniques from classical genomics, i.e. from *de novo* assembly of isolate genomes. We built upon continuous methodological advances with our genomic assembler Minia, adding new data structures such as the minimal perfect hash function [26] and the compressed graph representation. We participated in 2015 in the CAMI metagenomic reconstruction challenge [0]. This challenged gathered a total of 17 international groups, and Minia performed among the top assembly methods. This result is reported in an article to appear in Nature Methods from the CAMI consortium [25]. We further presented a poster at RECOMB 2017.

Targeted metagenomics. Within the PhD thesis of L. Siegwald, we have participated to the design of a comprehensive evaluation protocol to compare computational pipelines to analyze 16S amplicons, and have studied the impact of different variables on the biological interpretation of results. This study included the following tools: CLARK, Kraken, Mothur, Qiime and One Codex. It has been the subject of an invited keynote at the international workshop Recent Computational Advances in Metagenomics (RCAM 2017) [0].

7.2. Nonribosomal peptides

We further investigate the NRPs produced by *Burkholderia*, focusing on the identification of new compounds implicated in biocontrol and pharmaceutical [19].

New functionalities have been added to Norine to query the SMILES field either by the query form and the REST service. We also continue our curating of Norine data by improving peptide annotations and validation submissions of new peptides.

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

Researches on high-throughput V(D)J repertoire analysis started in the group in 2012. We have developed Vidjil, a web platform dedicated to the analysis of lymphocyte populations. Starting from DNA sequences, uploaded by the user, Vidjil identifies and quantifies lymphocyte populations and provides an interactive visualization.

[0]Tests of MATAM at MEDIS (INRA-Université Clermont Auvergne) for gene capture, Labgem (Genoscope) where it is on tracks to be integrated into the PathoTRACK-MicroScope platform dedicated to the human intestinal microbiome, and the Australian Centre for Ancient DNA (University of Adelaïde) for oral microbiome research.

[0]SFBI: Société Française de Bioinformatique

[0]CAMI challenge: https://data.cami-challenge.org/

[0]RCAM 2017: http://maiage.jouy.inra.fr/?q=fr/rcam2017

Seven European hospitals are now using Vidjil for their daily clinical practice. This year we published our experience of the minimal residual disease follow-up for acute lymphoblastic leukemia using Vidjil [24]. This is a first step towards using high-throughput sequencing and Vidjil for all the follow-up of the patients. We also participated to a joint publication with the EuroClonality-NGS consortium (see below).

Finally, we are working on transferring activities on platform development and user support. After meetings with several partners, we selected the Inria Foundation. The VidjilNet consortium (http://www.vidjil.net) will be launched in January 2018 within the InriaSoft action of the Foundation and will hire two engineers. VidjilNet will first gather hematology labs of French hospitals working on diagnosis and follow-up of acute lymphoblastic leukemia, and will be then extended to labs working on other pathologies as well as foreign labs.

7.4. RNA-Seq software benchmarking

Plenty of methods have been devised to analyze RNA-Seq data. Due to this large choice, it is a difficult task to determine what software is the best suited for a given question. To help in solving this problem, with colleagues at IRMB and in the SeqOne start-up in Montpellier, we devised a flexible benchmarking pipeline [18].

This pipeline is intended to be flexible enough to deal either with simulated or real data and to evaluate software on many possible aspects (mapping, splice detection, fusion detection, variant calling, and also on the post-analysis aspects such as gene quantification).

7.5. RNA folding landscape

Kinetics is key to understand many phenomena involving RNAs, such as co-transcriptional folding and riboswitches. Exact out-of-equilibrium studies induce extreme computational demands, leading state-of-the-art methods to rely on approximated kinetics landscapes, obtained using sampling strategies that strive to generate the key landmarks of the landscape topology. However, such methods are impeded by a large level of redundancy within sampled sets. Such a redundancy is uninformative, and obfuscates important intermediate states, leading to an incomplete vision of RNA dynamics.

Within the context of ANR RNAlands, we introduced RNANR, a new set of algorithms for the exploration of RNA kinetics landscapes at the secondary structure level. RNANR considers locally optimal structures, a reduced set of RNA conformations, in order to focus its sampling on basins in the kinetic landscape. Along with an exhaustive enumeration, RNANR implements a novel non-redundant stochastic sampling, and offers a rich array of structural parameters. Our tests on both real and random RNAs reveal that RNANR allows to generate more unique structures in a given time than its competitors, and allows a deeper exploration of kinetics landscapes [27].

7.6. Large-scale sequencing data indexing

Petabytes of DNA and RNA sequencing data are currently stored in online databases. It is currently possible to access these databases in two ways: 1) metadata queries, such as organism, instrument type, etc, and 2) download raw data. Due to the sheer size of the data, the web servers do not offer the possibility to search for sequences inside datasets. Such an operation would be invaluable to biology investigators, for example to determine which experiments contain an organism of interest, high expression of a certain transcript, a certain mutation, etc. Prior work exists for indexing sequencing data (Bloom Filter Tries, Sequence Bloom Trees), yet the performance remains prohibitive (either high memory usage, or several days for performing certain queries).

We proposed a new formalism, the Allsome Sequence Bloom Trees [28]. It improves upon Sequence Bloom Trees in terms of construction time (by 50%) and query time (by 40-85%), and also permits dataset-vs-dataset searches. The method has been tested by indexing a subset of 2,652 RNA-seq human experiments from the Sequence Read Archive. Allsome Sequence Bloom Trees pave the way towards "Google" searches of petabytes of sequencing data.

FUN Project-Team

7. New Results

7.1. Routing

Participants: Nathalie Mitton, Julien Vandaele.

Wireless sensor and actuator/robot networks need some routing mechanisms to ensure that data travel the network to the sink with some guarantees. The FUN research group has investigated different routing paradigms.

Geographic routing has gained much attention as a basic routing primitive in wireless sensor networks due to its memory-less, scalability, efficiency and low overhead features. Greedy forwarding is the simplest geographic routing scheme, it uses the distance as a forwarding criterion. Nevertheless, it may suffer from communication holes, where no next hop candidate is closer to the destination than the node currently holding the packet. For this purpose, a void handling technique is needed to recover from the void problem and successfully deliver data packets if a path does exist between source and destination nodes. Many approaches have been reported to solve this issue at the expense of extra processing and or overhead. [19] proposes GRACO, an efficient geographic routing protocol with a novel void recovery strategy based on ant colony optimization (ACO). GRACO is able to adaptively adjust the forwarding mechanism to avoid the blocking situation and effectively deliver data packets. Compared to GFG, one of the best performing geographic routing protocols, simulation results demonstrate that GRACO can successfully find shorter routing paths with higher delivery rate, less control packet overhead and shorter end-to-end delay.

Betweenness centrality metrics usually underestimate the importance of nodes that are close to shortest paths but do not exactly fall on them. In [16], [41], we reevaluate the importance of such nodes and propose the ρ-geodesic betweenness centrality, a novel metric that assigns weights to paths (and, consequently, to nodes on these paths) according to how close they are to shortest paths. The paths that are just slightly longer than the shortest one are defined as quasi-shortest paths, and they are able to increase or to decrease the importance of a node according to how often the node falls on them. We compare the proposed metric with the traditional, distance-scaled, and random walk betweenness centralities using four network datasets with distinct characteristics. The results show that the proposed metric, besides better assessing the topological role of a node, is also able to maintain the rank position of nodes overtime compared to the other metrics; this means that network dynamics affect less our metric than others. Such a property could help avoid, for instance, the waste of resources caused when data follow only the shortest paths and reduce associated costs.

To illustrate the data routing over a real demo, in [39], we show a webcam view of the testbed with remotely controlled lighting (ceiling LEDs and a mobile robot carrying a torch). A tight grid of 256 sensors will be used to collect light information. We display live updates of the resulting heatmap, live energy profiles and other performance metrics.

7.2. Security, Safety and Verification

Participants: Nathalie Mitton, Allan Blanchard, Simon Duquennoy.

Current practices of fault-tolerant network design ignore the fact that most network infrastructure faults are localized or spatially correlated (i.e., confined to geo-graphic regions). Network operators require new tools to mitigate the impact of such region-based faults on their infrastructures. Utilizing the support from the U.S. Department of Defense, and by consolidating a wide range of theories and solutions developed in the last few years, [12] designs RAPTOR, an advanced Network Planning and Management Tool that facilitates the design and provisioning of robust and resilient networks.The tool provides multi-faceted network design, evaluation, and simulation capabilities for network planners. Future extensions of the tool currently being worked upon not only expand the tool's capabilities, but also extend these capabilities to heterogeneous interdependent networks such as communication, power, water, and satellite networks.

IoT applications often utilize the cloud to store and provide ubiquitous access to collected data. This naturally facilitates data sharing with third-party services and other users, but bears privacy risks, due to data breaches or unauthorized trades with user data. To address these concerns, we present Pilatus, a data protection platform where the cloud stores only encrypted data, yet is still able to process certain queries (e.g., range, sum). More importantly, Pilatus features a novel encrypted data sharing scheme based on re-encryption, with revocation capabilities and in situ key-update. The solution proposed in [37], [56] includes a suite of novel techniques that enable efficient partially homomorphic encryption, decryption, and sharing. We present performance optimizations that render these cryptographic tools practical for mobile platforms. We implement a prototype of Pilatus and evaluate it thoroughly. Our optimizations achieve a performance gain within one order of magnitude compared to state-of-the-art realizations; mobile devices can decrypt hundreds of data points in a few hundred milliseconds. Moreover, we discuss practical considerations through two example mobile applications (Fitbit and Ava) that run Pilatus on real-world data.

7.3. Alternative communication paradigms

Participants: Antonio Costanzo, Valeria Loscri.

Nowadays, the always growing of connected objects and the strong demand to downsizing the devices in order to make the Internet of Things (IoT) paradigm more pervasive and ubiquitous, has motivated academic and industry people to investigate from one side mechanisms able to adapt quickly to the rapid external changes and to the quality of Services (QoS) parameters defined by the users and imposed by the adoption of new services and from another side, the investigation of portion of spectrum that have not been considered till this moment such as Terahertz band.

Bearing that in mind, we envisaged the possibility to leverage in a synergic way the Software Defined Radio (SDR) paradigm and the controlled mobility of mobiles wireless devices in order to adopt the most suitable modulation scheme and the best position with the objective to improve the network connectivity and coverage area [13].

On the other hand, spectrum scarcity and growing demand of nanocommunication systems have motivated researchers to investigation novel channel models in different portions of spectrum, namely in the THz band.

The fervent research activity in this direction is also motivated by the recent technological advances in new types of materials (e.g. graphene, novel metamaterials) presenting specific features suitable for this frequency spectrum and for the growing demand of downsizing antenna dimension.

In [15], we have investigated the chirality effect and Giant Optical Activity (GOA) and their impact when assuming different power allocation techniques.

On the other hand, when the nature of the matter and the interactions of specific particles and (quasi)particles such as phonons and photons are considered, there is a growing interest to investigate alternative communication paradigms based on these specific phenomena. In [14] we have performed an information theory analysis based on the generation of phonons elements when a source power as a cellphone is applied on biological tissue. The lesson learnt in this works is based on the consideration that where is heat transport it is possible to associate a communication paradigm. Follow this reasoning, in [50] we have revised the most recent advancement in terms of Visible Light Communication (VLC). Specifically, we have investigated Software Defined paradigm for VLC, in order to sketch out the main research directions for this new research domain.

7.4. Self-Organization

Participants: Nathalie Mitton, Valeria Loscri, Farouk Mezghani, Simon Duquennoy, Anjalalaina Jean Cristanel Razafimandimby.

7.4.1. Bayesian communications

In the last few years, Internet has become a very important vector of information sharing. Beyond the interconnection of computers and devices, there is still an important expansion capability, thanks to the capacity to interconnect heterogeneous devices. This extension of Internet known as Internet of Things (IoT) leads to (inter)connection of billions of objects. Nevertheless, IoT paradigm raises many challenges, such as the need to manage a massive amount of data generated by sensing devices. It was observed that, with the increase of sensors density, the redundancy of data increases. Thus, uploading raw data to the cloud can become extremely inefficient.

In order to address this issue, we proposed a Bayesian Inference Approach (BIA), able to remove a great amount of spatio-temporal correlated data [46], [10], [35].

In order to validate these approaches it was considered that experiments in real-world scenarios were needed. More specifically, we considered indoor tests in [46] and agricultural/outdoor experiments in [47]

7.4.2. Alert diffusion

Opportunistic communications present a promising solution as a disaster network recovery in emergency situations such as hurricanes, earthquakes and floods where infrastructure might be damaged. Recent works have proposed opportunistic-based alert diffusion approaches useful for trapped survivors. However, two main features were left behind. On the one hand, these works do not consider the assortment of networks integrated in mobile devices (e.g. WiFi-Direct, WiFi ad-hoc, cellular, bluetooth) and the choice of the network interface is left to the user who has no idea what is best or might be in a physical or psychological distress that impede the efficient selection. On the other hand, most of these works are based on selfish diffusion which might drain quickly the battery power. Moreover, they do not consider various energy level, which obviously influences the alert diffusion scheme. [17], [27], [28], [44] propose COPE and its demo, a cooperative opportunistic alert diffusion approach for disaster scenario that considers mobile devices that come with multiple network interfaces and with various battery power level. In order to maintain mobile devices alive longer, survivors form cliques and zones in which they diffuse alternately and periodically alert messages until reaching a potential rescuers team. Simulation results show that COPE largely outperforms the selfish diffusion scheme in terms of energy consumption while guaranteeing an important alert delivery success.

7.4.3. Consensus-based Leader election

In low-power wireless networking, new applications such as cooperative robots or industrial closed-loop control demand for network-wide consensus at low-latency and high reliability. Distributed consensus protocols is a mature eld of research in a wired context, but has received little attention in low-power wireless settings. In [21], [55], we present A^2: Agreement in the Air, a system that brings distributed consensus to low-power multi-hop networks. A^2 introduces Synchrotron, a synchronous transmissions kernel that builds a robust mesh by exploiting the capture effect, frequency hopping with parallel channels, and link-layer security. A^2 builds on top of this reliable base layer and enables the two-and three-phase commit protocols, as well as network services such as group membership, hopping sequence distribution and re-keying. We evaluate A^2 on four public testbeds with different deployment densities and sizes. A^2 requires only 475 ms to complete a two-phase commit over 180 nodes. The resulting duty cycle is 0.5% for 1-minute intervals. We show that A^2 achieves zero losses end-to-end over long experiments, representing millions of data points. When adding controlled failures, we show that two-phase commit ensures transaction consistency in A^2 while three-phase commit provides liveness at the expense of inconsistency under specific failure scenarios.

7.5. Smart Cities

Participants: Nathalie Mitton, Valeria Loscri, Abdoul Aziz Mbacke.

Smart cities are a key factor in the consumption of materials and resources. As populations grow and resources become scarcer, the efficient usage of these limited goods becomes more important. Building on and integrating with a huge amount of data, the cities of the future are becoming a realization today. There are millions of sensors in place already, monitoring various things in metropolises. In the near future, these sensors will multiply until they can monitor everything from streetlights and trashcans to road conditions and energy consumption. In this context, effective strategies or solutions for refining data sets can play a key role. Based on these premises, we propose in [32] intelligent and adaptive filtering mechanisms as a service (FIlAAS) integrated in the VITAL-OS middleware and will show their feasibility and their effectiveness in the smart city context.

Connecting all these devices to a cloud encompasses the execution of many network tasks at the *edge* and in particular on constrained gateways by low computational resources capabilities. Moreover, these gateways have to deal with the plethora of disparate technologies available in the IoT landscape. To cope with these issues, we introduce a Lightweight Edge Gateway for the Internet of Things (LEGIoT) architecture [18]. It relies on the modular characteristic of microservices and the flexibility of lightweight virtualization technologies to guarantee an extensible and flexible solution. In particular, by combining the implementation of specific frameworks and the benefits of container-based virtualization, our proposal enhances the suitability of edge gateways towards a wide variety of IoT protocols/applications (for both downlink and uplink) enabling an optimized resource management and taking into account requirements such as energy efficiency, multi-tenancy, and interoperability. LEGIoT is designed to be hardware agnostic and its implementation has been tested within a real sensor network. Achieved results demonstrate its scalability and suitability to host different applications meant to provide a wide range of IoT services.

In parallel, we proposed a MOOC in the framework of the IPL CityLab project (See Section 9.2.1), whose working documents are available online [51], [52], [53], [54].

7.6. Smart Grids

Participants: Nathalie Mitton, Jad Nassar.

The Internet of Thing is a on going revolution which promises to interconnect most of our world with billions of connected devices. Hence, data routing and prioritization in IoT is a main challenge in this gigantic network. This is all the more true for the Smart Grids data management where heterogeneous applications and signaling messages have different requirements in terms of reliability, latency and priority. So far, standards on Smart Grid have recommended the use of RPL (Routing Protocol for Low-Power and Lossy networks) protocol for distributing commands over the grid. RPL assures Quality of Service (QoS) at the network layer in wireless sensor networks through the logical subdivision of the network in multiple instances, each one relying on a specific Objective Function. However, RPL is not optimized for Smart Grids, as its main objective function and its associated metric does not allow for QoS differentiation. In order to overcome this, in [31], [45] we propose *OFQS* an objective function with a multi-objective metric that considers the delay and the remaining energy in the battery nodes alongside with the quality of the links. Our function automatically adapts to the number of instances (traffic classes) providing a QoS differentiation based on the different Smart Grid applications requirements. Simulations show that our proposal provides a low packet delivery latency and a higher packet delivery ratio while extending the lifetime of the network compared to literature solutions.

In the same spirit, we have proposed QoSGRACO [36], a routing protocol which takes account of the Quality of Service (QoS) of NAN's traffic by using colored pheromones ant colonies. We show, through simulations, that QoS-GRACO is able to satisfy NAN's requirements, especially in terms of delay and reliability.

7.7. Vehicular networks and smart car platforms

Participants: Nathalie Mitton, Valeria Loscri.

Roadside Units (RSUs) are an important component of vehicular ad hoc networks (VANETs). In a VANET, RSUs are deployed at intersections or some points along a road to help improve network connectivity, data delivery, and thus network services to vehicles. Therefore, RSU deployment has a big impact on the network performance and becomes an important issue in the design of a VANET. In general, RSU deployment is costly. To achieve a good tradeoff between deployment cost and network performance, it is expected to optimize the deployment of RSUs in a VANET. To address this challenge, considerable research work has been conducted on the optimization of RSU deployment for VANET.

To optimize the RSU deployment, the notion of centrality in a social network to RSU deployment has been introduced [40], and use it to measure the importance of an RSU position candidate in RSU deployment. Based on the notion of centrality, we propose a centrality-based RSU deployment approach and formulate the RSU deployment problem as a linear programing problem with the objective to maximize the total centrality of all position candidates selected for RSU deployment under the constraint of a given deployment budget.

Nowadays, many vehicular applications are emerging, arising from entertainment and road safety. The paradigm of Internet of Vehicles (IOV) is proliferating and it is an inevitable convergence of the existing mobile Internet and the concept of Internet-of-Things (IoT). In IoV, Internet of applications go on "wheels", then converting the existing vehicles into "digital cars" equipped with several technologies and sensors. The meeting of mobility with social interactions rises to a particular class of social networks, Vehicular Social Networks [20]. Vehicular social networks are online social networks where the social interactions are built *on-the-fly*, due to the opportunistic links in vehicular networks. In this context, the reliability of message dissemination becomes very important and the adding of social components allows the definition of trust parameter to be directly included in the forwarding technique.

On the other hand, the huge amount of sensors deployed in a car, are arising new types of applications and services, even if data are "confined" in the car.

However, with the increasing number of functionalities, computing and communication resources, that have to be handled by the On Board Unit (OBU), is constantly growing. OBUs are embedded systems with limited hardware resources and a critical software design that makes also a simple update procedure a not trivial operation. These constraints are combined with the typical software lifetime, which is much shorter than the lifetime of mechanical and technological components. Bearing all these points in mind, we have identified in the use of lightweight virtualization technologies a suitable mechanism, which allows to design an OBU that can satisfy several requirements in terms of efficient software and hard- ware resources management [29], [30].

7.8. Robots

Participants: Nathalie Mitton, Valeria Loscri, Anjalalaina Jean Cristanel Razafimandimby.

With the advent of the Internet of Things (IoT) and robotics on one hand and of Cloud Computing on the other hand, people have witnessed a shift in the way they can interact and communicate with their things and their environment. Together with these main concepts, the vision of Robot as a Service can be considered and applied to different contexts and domains. Unfortunately, Remote programming and control of heterogeneous robots are not always possible, which ends up as difficult tasks requiring advanced skills. in order to face these challenges, Open Mobile Cloud Robotics Interface (OMCRI) has been proposed in [26]. It represents an extension of OCCI platform based on the Robot-as-Service paradigm. OMCRI presents interesting features such as modularity and extensibility.

Another interesting concept recently introduced by ABY Research, is the Internet of Robotic Things (IoRT), representing a dynamic actuation. In [48], to realize an efficient deployment and an effective coverage by also keeping a good communication quality, we have proposed an IoT- based and neural network control scheme. The neural network controller, in turn, is completely distributed and mimics perfectly the IoT-based approach. Results show that our approaches are efficient, in terms of convergence time, connectivity, and energy consumption.

7.9. MAC mechanisms

Participants: Nathalie Mitton, Simon Duquennoy, Viktor Toldov.

In the era of the Internet of Things (IoT), the number of connected devices is growing dramatically. Often, connected objects use Industrial, Scientific and Medical (ISM) radio bands for communication. These kinds of bands are available without license, which facilitates development and implementation of new connected objects. However, it also leads to an increased level of interference in these bands. Interference not only negatively affect the Quality of Service, but also causes energy losses, which is especially unfavorable for the energy constrained Wireless Sensor Networks (WSN). In [11], the impact of the interference on the energy consumption of the WSN nodes is studied experimentally. The experimental results were used to estimate the lifetime of WSN nodes under conditions of different levels of interference. Then, a Thompson sampling based Cognitive Radio adaptive solution is proposed and evaluated via both, simulation and hardware implementation. Results show that this approach finds the best channel quicker than other state of the art solutions. Based on a set of experimentations, an adaptive WildMAC MAC layer protocol is proposed and evaluated experimentally.

In parallel, synchronized communication has recently emerged as a prime option for low-power critical applications. Solutions such as Glossy or Time Slotted Channel Hopping (TSCH) have demonstrated end-to-end reliability upwards of 99.9%. In this context, the IETF Working Group 6TiSCH is currently standardizing the mechanisms to use TSCH in low-power IPv6 scenarios [42] identifies a number of challenges when it comes to implementing the 6TiSCH stack [43]. It shows how these challenges can be addressed with practical solutions for locking, queuing, scheduling and other aspects. With this implementation as an enabler, we present an experimental validation and comparison with state-of-the-art MAC protocols. We conduct fine-grained energy profiling, showing the impact of link-layer security on packet transmission. We evaluate distributed time synchronization in a 340-node testbed, and demonstrate that tight synchronization (hundreds of microseconds) can be achieved at very low cost (0.3% duty cycle, 0.008% channel utilization). We finally compare TSCH against traditional MAC layers: low-power listening (LPL) and CSMA, in terms of reliability, latency and energy. We show that with proper scheduling, TSCH achieves by far the highest reliability, and outperforms LPL in both energy and latency.

7.10. RFID

Participants: Nathalie Mitton, Abdoul Aziz Mbacke, Ibrahim Amadou, Gabriele Sabatino.

The advent of RFID (Radio Frequency Identification) has allowed the development of numerous applications. Indeed, solutions such as tracking of goods in large areas or sensing in smart cities are now made possible. However, such solutions encounter two main issues, first is inherent to the technology itself which is readers collisions, the second one being the gathering of read data up to a base station, potentially in a multihop fashion. While the first one has been a main research subject in the late years, the next one has not been investigated for the sole purpose of RFID, but rather for wireless adhoc networks. This multihop tag information collection must be done in regards of the application requirements but it should also care for the deployment strategy of readers to take advantage of their relative positions, coverage, reading activity and deployment density to avoid interfering between tag reading and data forwarding. To the best of our knowledge, the issue for a joint scheduling between tag reading and forwarding has never been investigated so far in the literature, although important. [24] addresses the anti-collision issue in mobile environments. In [23], we propose two new distributed, crosslayer solutions meant for the reduction of collisions and better efficiency of the RFID system but also serve as a routing solution towards a base station. Simulations show high levels of throughput while not lowering on the fairness on medium access staying above 85% in the highest deployment density with up to 500 readers, also providing a 90% data rate. In [25], we propose two distributed and efficient solutions for dense mobile deployments of RFID systems. mDEFAR is an adaptation of a previous work highly performing in terms of collisions reduction, efficiency and fairness in dense static deployments. CORA is more of a locally mutual solution where each reader relies on its neighborhood to enable itself or not. Using a beaconing mechanism, each reader is able to identify potential (non-)colliding neighbors in a running frame and as such chooses to read or not. Performance evaluation shows high performance in terms

of coverage delay for both proposals quickly achieving 100% coverage depending on the considered use case while always maintaining consistent efficiency levels above 70%. Compared to GDRA, our solutions proved to be better suited for highly dense and mobile environments, offering both higher throughput and efficiency. The results reveal that depending on the application considered, choosing either mDEFAR or CORA helps improve efficiency and coverage delay.

RMOD Project-Team

7. New Results

7.1. Software Quality: Testing and Tools

Testing Habits. What are the Testing Habits of Developers? We conducted a case study in a large IT company. Tests are considered important to ensure the good behavior of applications and improve their quality. But development in companies also involves tight schedules, old habits, less-trained developers, or practical difficulties such as creating a test database. As a result, good testing practices are not always used as often as one might wish. With a major IT company, we are engaged in a project to understand developers testing behavior, and whether it can be improved. Some ideas are to promote testing by reducing test session length, or by running automatically tests behind the scene and send warnings to developers about the failing ones. Reports on developers testing habits in the literature focus on highly distributed open-source projects, or involve students programmers. As such they might not apply to our industrial, closed source, context. We take inspiration from experiments of two papers of the literature to enhance our comprehension of the industrial environment. We report the results of a field study on how often the developers use tests in their daily practice, whether they make use of tests selection and why they do. Results are reinforced by interviews with developers involved in the study. The main findings are that test practice is in better shape than we expected; developers select tests ruthlessly (instead of launching an entire test suite); although they are not accurate in their selection, and; contrary to expectation, test selection is not influenced by the size of the test suite nor the duration of the tests. [23]

Tests in Open-Source. During the development, it is known that tests ensure the good behavior of applications and improve their quality. We studied developers testing behavior inside the Pharo community in the purpose to improve it. We report results of a field study on how often the developers use tests in their daily practice, whether they make use of tests selection and why they do. Results are strengthened by interviews with developers involved in the study. The main findings are that developers run tests every modifications of their code they did; most of the time they practice test selection (instead of launching an entire test suite); however they are not accurate in their selection; they change their selection depending on the duration of the tests and; contrary to expectation, test selection is not influenced by the size of the test suite. [35]

CodeCritics Applied to Database Schema: Challenges and First Results. Relational databases (DB) play a critical role in many information systems. For different reasons, their schemas gather not only tables and columns but also views, triggers or stored functions (i.e., fragments of code describing treatments). As for any other code-related artefact, software quality in a DB schema helps avoiding future bugs. However, few tools exist to analyze DB quality and prevent the introduction of technical debt. We present research issues related to assessing the software quality of a DB schema by adapting existing source code analysis research to database schemas. We present preliminary results that have been validated through the implementation of DBCritics, a prototype tool to perform static analysis on the SQL source code of a database schema. DBCritics addresses the limitations of existing DB quality tools based on an internal representation considering all entities of the database and their relationships. [26]

Recommending Source Code Locations for System Specific Transformations. From time to time, developers perform sequences of code transformations in a systematic and repetitive way. This may happen, for example, when introducing a design pattern in a legacy system: similar classes have to be introduced, containing similar methods that are called in a similar way. Automation of these sequences of transformations has been proposed in the literature to avoid errors due to their repetitive nature. However, developers still need support to identify all the relevant code locations that are candidate for transformation. Past research showed that these kinds of transformation can lag for years with forgotten instances popping out from time to time as other evolutions bring them into light. We evaluate three distinct code search approaches (structural, based on Information Retrieval, and AST based algorithm) to find code locations that would require similar transformations. We validate the resulting candidate locations from these approaches on real cases identified previously

in literature. The results show that looking for code with similar roles, e.g., classes in the same hierarchy, provides interesting results with an average recall of 87% and in some cases the precision up to 70%. [33]

Quality-oriented Move Method Refactoring. Restructuring is an important activity to improve software internal structure. Even though there are many restructuring approaches, very few consider the refactoring impact on the software quality. In this paper, we propose an semi-automatic software restructuring approach based on quality attributes. We rely on the measurements of the Quality Model for Object Oriented Design (QMOOD) to recommend Move Method refactorings that improve software quality. In our preliminary evaluation on three open-source systems, our approach achieved an average recall of 57%. [34]

7.2. Software Reengineering

The Case for Non-Cohesive Packages. While the lack of cohesiveness of modules in procedural languages is a good way to identify modules with potential quality problems, we doubt that it is an adequate measure for packages in object-oriented systems. Indeed, mapping procedural metrics to object-oriented systems should take into account the building principles of object-oriented programming: inheritance and late binding. Inheritance offers the possibility to create packages by just extending classes with the necessary increment of behavior. Late binding coupled to the "Hollywood Principle" are a key to build frameworks and let the users branch their extensions in the framework. Therefore, a package extending a framework does not have to be cohesive, since it inherits the framework logic, which is encapsulated in framework packages. In such a case, the correct modularization of an extender application may imply low cohesion for some of the packages. We confirm these conjectures on various real systems (JHotdraw, Eclipse, JEdit, JFace) using or extending OO frameworks. We carry out a dependency analysis of packages to measure their relation with their framework. The results show that framework dependencies form a considerable portion of the overall package dependencies. This means that non-cohesive packages should not be considered systematically as packages of low quality. [22]

Identifying Classes in Legacy JavaScript Code. JavaScript is the most popular programming language for the Web. Although the language is prototype-based, developers can emulate class-based abstractions in JavaScript to master the increasing complexity of their applications. Identifying classes in legacy JavaScript code can support these developers at least in the following activities: (i) program comprehension; (ii) migration to the new JavaScript syntax that supports classes; and (iii) implementation of supporting tools, including IDEs with class-based views and reverse engineering tools. We propose a strategy to detect class-based abstractions in the source code of legacy JavaScript systems. We report on a large and in-depth study to understand how class emulation is employed, using a dataset of 918 JavaScript applications available on GitHub. We found that almost 70% of the JavaScript systems we study make some usage of classes. We also performed a field study with the main developers of 60 popular JavaScript systems in order to validate our findings. The overall results range from 97% to 100% for precision, from 70% to 89% for recall, and from 82% to 94% for F-score. [20]

A Critical Analysis of String APIs: The Case of Pharo. Most programming languages, besides C, provide a native abstraction for character strings, but string APIs vary widely in size, expressiveness, and subjective convenience across languages. In Pharo, while at first glance the API of the String class seems rich, it often feels cumbersome in practice; to improve its usability, we faced the challenge of assessing its design. However, we found hardly any guideline about design forces and how they structure the design space, and no comprehensive analysis of the expected string operations and their different variations. We first analyze the Pharo 4 String library, then contrast it with its Haskell, Java, Python, Ruby, and Rust counterparts. We harvest criteria to describe a string API, and reflect on features and design tensions. This analysis should help language designers in understanding the design space of strings, and will serve as a basis for a future redesign of the string library in Pharo. [19]

7.3. Dynamic Languages: Language Constructs for Modular Design

An Experiment with Lexically-bound Extension Methods for a Dynamic Language. An extension method is a method declared in a package other than the package of its host class. Thanks to extension methods,

developers can adapt classes they do not own to their needs: adding methods to core classes is a typical use case. This is particularly useful for adapting software and therefore increasing reusability. In most dynamically-typed languages, extension methods are globally visible. Because any developer can define extension methods for any class, naming conflicts occur: if two developers define an extension method with the same signature in the same class, only one extension method is visible and overwrites the other. Similarly, if two developers each define an extension method with the same name in a class hierarchy, one overrides the other. Existing solutions typically rely on a dedicated and slow method lookup algorithm to resolve conflicts at runtime. We present a model of scoped extension methods that minimizes accidental overrides and we present an implementation in Pharo that incurs little performance overhead. This implementation is based on lexical scope and hierarchy-first strategy for extension scoping. [44]

Scoped Extension Methods in Dynamically-Typed Languages. An extension method is a method declared in a package other than the package of its host class. Thanks to extension methods, developers can adapt to their needs classes they do not own: adding methods to core classes is a typical use case. This is particularly useful for adapting software and therefore to increase reusability. Inquiry. In most dynamically-typed languages, extension methods are globally visible. Because any developer can define extension methods for any class, naming conflicts occur: if two developers define an extension method with the same signature in the same class, only one extension method is visible and overwrites the other. Similarly, if two developers each define an extension method with the same name in a class hierarchy, one overrides the other. To avoid such *accidental overrides*, some dynamically-typed languages limit the visibility of an extension method to a particular scope. However, their semantics have not been fully described and compared. In addition, these solutions typically rely on a dedicated and slow method lookup algorithm to resolve conflicts at runtime. Approach. In this article, we present a formalization of the underlying models of Ruby refinements, Groovy categories, Classboxes, and Method Shelters that are scoping extension method solutions in dynamically-typed languages. Knowledge. Our formal framework allows us to objectively compare and analyze the shortcomings of the studied solutions and other different approaches such as MultiJava. In addition, language designers can use our formal framework to determine which mechanism has less risk of *accidental overrides*. Grounding. Our comparison and analysis of existing solutions is grounded because it is based on denotational semantics formalizations. Importance. Extension methods are widely used in programming languages that support them, especially dynamically-typed languages such as Pharo, Ruby or Python. However, without a carefully designed mechanism, this feature can cause insidious hidden bugs or can be voluntarily used to gain access to protected operations, violate encapsulation or break fundamental invariants. [17]

First-Class Undefined Classes for Pharo: From Alternative Designs to a Unified Practical Solution. Loading code inside a Pharo image is a daily concern for a Pharo developer. Nevertheless, several problems may arise at loading time that can prevent the code to load or even worse let the system in an inconsistent state. We focus on the problem of loading code that references a class that does not exist in the system. We discuss the different flavors of this problem, the limitations of the existing Undeclared mechanism and the heterogeneity of Pharo tools to solve it. Then, we propose an unified solution for Pharo that reifies Undefined Classes. Our model of Undefined Classes is the result of an objective selection among different alternatives. We then validate our solution through two cases studies: migrating old code and loading code with circular dependencies. We also present the integration of this solution into Pharo regarding the needed Meta-Object Protocol for Undefined Classes and the required modifications of existing tools. [30]

Run-Fail-Grow: Creating Tailored Object-Oriented Runtimes. Producing a small deployment version of an application is a challenge because static abstractions such as packages cannot anticipate the use of their parts at runtime. Thus, an application often occupies more memory than actually needed. Tailoring is one of the main solutions to this problem i.e., extracting used code units such as classes and methods of an application. However, existing tailoring techniques are mostly based on static type annotations. These techniques cannot efficiently tailor applications in all their extent (e.g., runtime object graphs and metadata) nor be used in the context of dynamically-typed languages. We propose a run-fail-grow technique to tailor applications using their runtime execution. Run-fail-grow launches (a) a reference application containing the original application to tailor and (b) a nurtured application containing only a seed with a minimal set of code units the user wants to ensure in the final application. The nurtured application is executed, failing when it founds missing objects,

classes or methods. On failure, the necessary elements are installed into the nurtured application from the reference one, and the execution resumes. The nurtured application is executed until it finishes, or until the developer explicitly finishes it, for example in the case of a web application. resulting in an object memory (i.e., a heap) with only objects, classes and methods required to execute the application. To validate our approach we implemented a tool based on Virtual Machine modifications, namely Tornado. Tornado succeeds to create very small memory footprint versions of applications e.g., a simple object-oriented heap of 11kb. We show how tailoring works on application code, base and third-party libraries even supporting human interaction with user G. interfaces. These experiments show memory savings ranging from 95% to 99%. [18]

7.4. Dynamic Languages: Debugging

Unanticipated Debugging with Dynamic Layers. To debug running software we need unanticipated adaptation capabilities, especially when systems cannot be stopped, updated and restarted. Adapting such programs at runtime is an extreme solution given the delicate live contexts the debugging activity takes place. We introduce the Dynamic Layer, a construct in which behavioral variations are gathered and activated as a whole set of adaptations. Dimensions of Dynamic Layers activation are reified to allow very fine definitions of layer scopes and a fine grained selection of adapted entities. An experimental implementation with the Pharo language is evaluated through a runtime adaptation example. [25]

New Generation Debuggers. Locating and fixing bugs is a well-known time consuming task. Advanced approaches such as object-centric or back-in-time debuggers have been proposed in the literature, still in many scenarios developers are left alone with primitive tools such as manual breakpoints and execution stepping. We explore several advanced on-line debugging techniques such as advanced breakpoints and on-line execution comparison, that could help developers solve complex debugging scenarios. We analyze the challenges and underlying mechanisms required by these techniques. We present some early but promising prototypes we built on the Pharo programming language. We finally identify future research paths by analyzing existing research and connecting it to the techniques we presented before. [27]

Debugging Cyber-Physical Systems. Cyber-Physical Systems (CPS) integrate sensors and actuators to collect data and control entities in the physical world. Debugging CPS systems is hard due to the time-sensitive nature of a distributed applications combined with the lack of control on the surrounding physical environment. This makes bugs in CPS systems hard to reproduce and thus to fix. In this context, on-line debugging techniques are helpful because the debugger is connected to the device when an exception or crash occurs. We report on our experiences on applying two different on-line debugging techniques for a CPS system: remote debugging using the Pharo remote debugger and our IDRA debugger. In contrast to traditional remote debugging, IDRA allows to on-line debug an application locally in another client machine by reproducing the runtime context where the bug manifested. Our qualitative evaluation shows that IDRA provides almost the same interaction capabilities than Pharo's remote debugger and is less intrusive when performing hot-modifications. Our benchmarks also show that IDRA is significantly faster than the Pharo remote debugger, although it increases the amount of data transferred over the network. [29]

Reflectogram. Reflective facilities in OO languages are used both for implementing language extensions (such as AOP frameworks) and for supporting new programming tools and methodologies (such as object-centric debugging and message-based profiling). Yet controlling the runtime behavior of these reflective facilities introduces several challenges, such as computational overhead, the possibility of meta-recursion and an unclean separation of concerns between base and meta-level. We present five dimensions of meta-level control from related literature that try to remedy these problems. These dimensions are namely: temporal and spatial control, placement control, level control and identity control. We then discuss how these dimensions interact with language semantics in class-based OO languages in terms of: scoping, inheritance and first-class entities. We argue that the reification of the descriptive notion of reflectogram can unify the control of meta-level execution in all these five dimensions while expressing properly the underlying language semantics. We present an extended model for the reification of the reflectogram based on our additional analysis and validate our approach through a new prototype implementation that relies on byte-code instrumentation. Finally, we illustrate our approach through a case study on runtime tracing. [16]

7.5. Dynamic Languages: Virtual Machines

VM Profiler. Code profiling enables a user to know where in an application or function the execution time is spent. The Pharo ecosystem offers several code profilers. However, most of the publicly available profilers (MessageTally, Spy, GadgetProfiler) largely ignore the activity carried out by the virtual machine, thus incurring inaccuracy in the gathered information and missing important information, such as the Just-in-time compiler activity. We describe the motivations and the latest improvements carried out in VMProfiler, a code execution pro-filer hooked into the virtual machine, that performs its analysis by monitoring the virtual machine execution. These improvements address some limitations related to assessing the activity of native functions (resulting from a Just-in-time compiler operation): as of now, VMProfiler provides more detailed profiling reports, showing for native code functions in which bytecode range the execution time is spent. [28]

Sista: Saving Optimized Code in Snapshots for Fast Start-Up. Modern virtual machines for object-oriented languages such as Java HotSpot, Javascript V8 or Python PyPy reach high performance through just-in-time compilation techniques, involving on-the-fly optimization and deoptimization of the executed code. These techniques require a warm-up time for the virtual machine to collect information about the code it executes to be able to generate highly optimized code. This warm-up time required before reaching peak performance can be considerable and problematic. We propose an approach, Sista (Speculative Inlining SmallTalk Architecture) to persist optimized code in a platform-independent representation as part of a snapshot. After explaining the overall approach, we show on a large set of benchmarks that the Sista virtual machine can reach peak performance almost immediately after start-up when using a snapshot where optimized code was persisted. [24]

7.6. Interaction

This work is done in collaboration with team Mjolnir.

Turning Function Calls Into Animations. Animated transitions are an integral part of modern interaction frameworks. With the increasing number of animation scenarios, they have grown in range of animatable features. Yet not all transitions can be smoothed: programming systems limit the flexibility of frameworks for animating new things, and force them to expose low-level details to programmers. We present an ongoing work to provide system-wide animation of objects, by introducing a delay operator. This operator turns setter function calls into animations. It offers a coherent way to express animations across frameworks, and facilitates the animation of new properties. [31]

7.7. Software Engineering for BlockChain and Smart Contracts

Solidity Parsing Using SmaCC: Challenges and Irregularities. Solidity is a language used to implement smart contracts on a blockchain platform. Since its initial conception in 2014, Solidity has evolved into one of the major languages for the Ethereum platform as well as other blockchain technologies. Due to its popularity, there are many tools specifically designed to handle smart contracts written in Solidity. However , there is a lack of tools for Pharo to handle Solidity contracts. Therefore, we implemented a parser using SmaCC to serve as a base for further developing Solidity support in Pharo. We describe the parser creation, the irregularities we found in the Solidity grammar specification, and common practices on how to adapt the grammar to an LR type parser. Our experiences with parsing the Solidity language using SmaCC may help other developers trying to convert similar grammars. [32]

SmartInspect: Smart Contract Inspection. Smart contracts are embedded procedures stored with the data they act upon. Debugging deployed Smart Contracts is a difficult task since once deployed, the code cannot be reexecuted and inspecting a simple attribute is not easily possible because data is encoded. In this technical report, we present SmartInspect to address the lack of inspectability of a deployed contract. Our solution analyses the contract state by using decompilation techniques and a mirror-based architecture to represent the object responsible for interpreting the contract state. SmartInspect allows developers and also end-users of a contract to better visualize and understand the contract stored state without needing to redeploy, nor develop any ad-hoc code. [43]

SPIRALS Project-Team

7. New Results

7.1. A Domain-specific Language for The Control of Self-adaptive Component-based Architecture

In [12], together with Frederico Alvares (Inria Ascola) and Eric Rutten (Inria Ctrl-A), we have proposed Ctrl-F, a new domain-specific language for specifying reconfiguration policies in self-adaptable component-based software systems. Self-adaptive behaviors in the context of component-based architecture are generally designed based on past monitoring events, configurations (component assemblies) as well as behavioral programs defining the adaptation logics and invariant properties. The novelty of the proposed Ctrl-F language is to enable taking decisions on predictions on the possible futures of the system in order to avoid going into branches of the behavioral program leading to bad configurations. Ctrl-F is formally defined by a translation into *Finite State Automata* models. We use *Discrete Controller Synthesis* to automatically generate a controller to enforce correct self-adaptive behaviors. Ctrl-F is integrated with our FraSCAti middleware platform for distributed service and component oriented systems.

7.2. A New Interface for Mobile Cloud Robotics

In [35], together with Nathalie Mitton (Inria Fun), we have proposed OMCRI, a new interface for mobile cloud robotics. This interface enables to abstract from the heterogeneity of robotic platforms and to bring some resource management facilities to fleets of robots. This result is based on the expertise that we have developed in the management of resources for cloud computing environments, especially around the OCCI standard. To the best of our knowledge, OMCRI is the first interface that enables to concretize the vision of robotics as a service. This result has obtained the best award at the 2nd IEEE International Congress on Internet of Things (ICIOT 2017).

DEFROST Project-Team

7. New Results

7.1. New results

7.1.1. Inverse model with contact handling

Publication in RA-Letter 2017 (proceeding ICRA 2017): Optimization-based inverse model of soft robots with contact handling, E. Coevoet, A. Escande, C. Duriez. Abstract. This paper presents a physically-based algorithm to interactively simulate and control the motion of soft robots interacting with their environment. We use the Finite Element Method (FEM) to simulate the non-linear deformation of the soft structure, its actuators, and surroundings, and propose a control method relying on a quadratic optimization to find the inverse of the model. The novelty of this work is that the deformations due to contacts, including self-collisions, are taken into account in the optimization process. We propose a dedicated and efficient solver to handle the linear complementarity constraints introduced by the contacts. Thus, the method allows interactive transfer of the motion of soft robots from their task space to their actuator space while interacting with their surrounding. The method is generic and tested on several numerical examples and on a real cable-driven soft robot. http://hal.univ-lille3.fr/hal-01500912

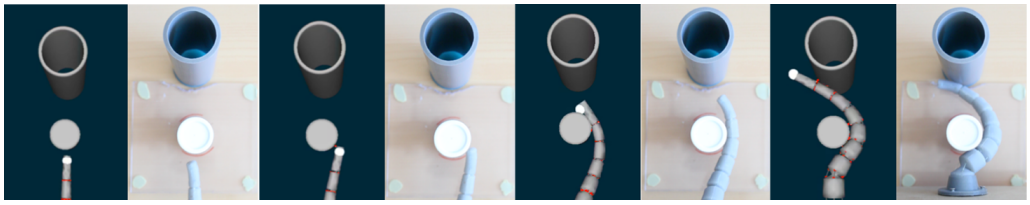

Figure 2. Top: Real cable-driven soft robot, Bottom: Corresponding motion computed by the simulated inverse model. The input of the inverse model is the motion of the robot's tip.

7.1.2. Hydraulic actuators

Publication at ICRA 2017: Real-time simulation of hydraulic components for interactive control of soft robots, A. Rodrìguez, E. Coevoet, C. Duriez. Abstract. In this work we propose a new method for online motion planning in the task-space for hydraulic actuated soft robots. Our solution relies on the interactive resolution of an inverse kinematics problem, that takes into account the properties (mass, stiffness) of the deformable material used to build the robot. An accurate modeling of the mechanical behavior of hydraulic components is based on a novel GPU parallel method for the real-time computation of fluid weight distribution. The efficiency of the method is further increased by a novel GPU parallel leveraging mechanism. Our complete solution has been integrated within the open-source SOFA framework. In our results, we validate our simulation with a fabricated silicone cylinder and we demonstrate the usage of our approach for direct control of hydraulic soft robots. https://hal.inria.fr/hal-01535810

7.1.3. Visual Servoing Control of Soft Robots

Publication at IROS 2017: Visual Servoing Control of Soft Robots based on Finite Element Model, Zhongkai Zhang, Thor Morales Bieze, Jeremie Dequidt, Alexandre Kruszewski, Christian Duriez. Abstract. In this paper, we propose a strategy for the control of soft robots with visual tracking and simulation-based predictor. A

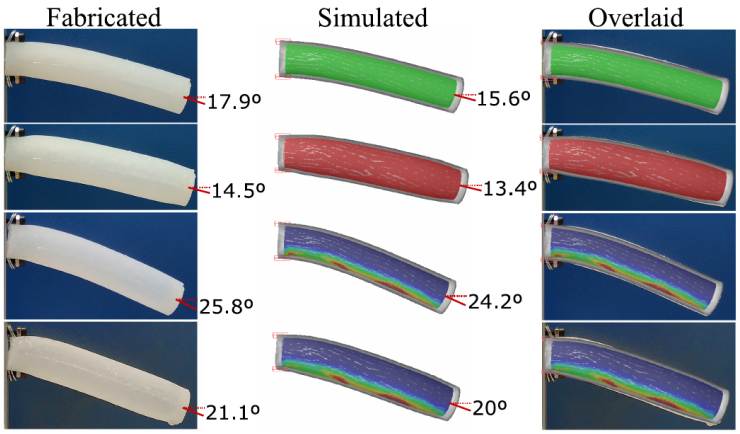

Figure 3. Comparison between experiments and simulation, with hydraulic or air actuation.

kinematic model of soft robots is obtained thanks to the Finite Element Method (FEM) computed in real-time. The FEM allows to obtain a prediction of the Jacobian matrix of the robot. This allows a first control of the robot, in the actuator space. Then, a second control strategy based on the feedback of infrared cameras is developed to obtain a correction of the effector position. The robust stability of this closed-loop system is obtained based on Lyapunov stability theory. Otherwise, to deal with the problem of image features (the marker points placed on the end effector of soft robot) loss, a switched control strategy is proposed to combine both the open-loop controller and the closed-loop controller. Finally, experiments on a parallel soft robot driven by four cables are conducted and show the effectiveness of these methods for the real-time control of soft robots.
https://hal.inria.fr/hal-01618330

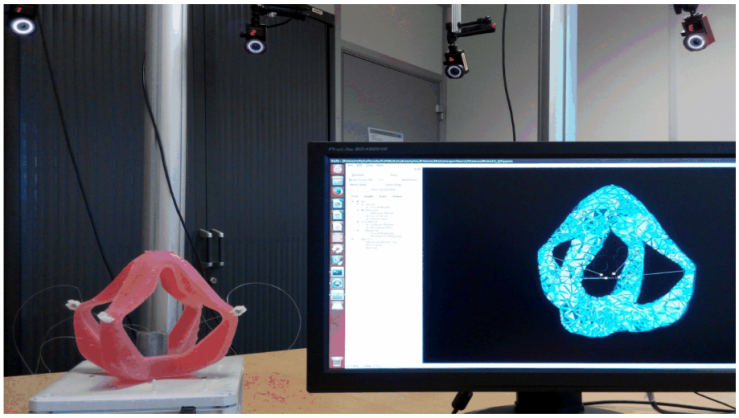

Figure 4. Experimental setup for trajectory tracking of the soft robot. The robot has four soft arms and is actuated by cables. The simulation model is used as a simulation-based predictor. The 3D position of the end effector can be obtained by the position perception system.

7.1.4. Dynamic control of soft robots

Simulation results have been presented at the International Federation of Automatic Control (IFAC) World Congress in Toulouse in July. We presented a control strategy build on a reduced order dynamical model of a soft robot and showed that it allows the user to make the studied soft robot converge faster to a desired position without oscillations. This paper was illustrated with simulation experiments. https://hal.archives-ouvertes.fr/hal-01558844/

7.1.5. Framework for soft robot simulation

Publication in Advanced Robotics 2017: Software toolkit for modeling, simulation and control of soft robots, E. Coevoet, T. Morales-Bieze, F. Largilliere, Z. Zhang, M. Thieffry, M. Sanz-Lopez, B. Carrez, D. Marchal, O. Goury, J. Dequidt, C. Duriez. Abstract. The technological differences between traditional robotics and soft robotics have an impact on all of the modeling tools generally in use, including direct kinematics and inverse models, Jacobians, and dynamics. Due to the lack of precise modeling and control methods for soft robots, the promising concepts of using such design for complex applications (medicine, assistance, domestic robotics...) cannot be practically implemented. This paper presents a first unified software framework dedicated to modeling, simulation and control of soft robots. The framework relies on continuum mechanics for modeling the robotic parts and boundary conditions like actuators or contacts using a unified representation based on Lagrange multipliers. It enables the digital robot to be simulated in its environment using a direct model. The model can also be inverted online using an optimization-based method which allows to control the physical robots in the task space. To demonstrate the effectiveness of the approach, we present various soft robots scenarios including ones where the robot is interacting with its environment. The software has been built on top of SOFA, an open-source framework for deformable online simulation and is available at https://project.inria.fr/softrobot/. https://hal.inria.fr/hal-01649355

7.1.6. Shape modelling and optimization

Thomas Morzadec and Erwan Douaille are working on new approaches to model the soft robots' shape. They are working under the supervision of Damien Marchal and Christian Duriez. During this year, they developped a modelling kernel based on distance field function and integrate it into Sofa. The geometrical representation offer several interesting properties that are: compactness and generative. For usability the distance functions are described using the python language then transpiled into more efficient languages (glsl and cython). This geometric kernel is planned for integration into Sofa. On top of this framework shape optimization are now investigated using genetic algorithms.

LINKS Project-Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Aggregates

Aggregation refers to the computation of aggregates in databases, that is, the computation of a function of the answer of a query, such as counting the number of answers, finding the optimal one for a given objective function or enumerating all of them with a small delay between two distinct answers. The goal of aggregation is typically to compute such aggregates without explicitly generating the whole set of answers. We study aggregation problem within the ANR project *Aggreg* coordinated by Niehren.

At ***ICALP*** Bourhis (with Amarilli, Jachiet and Mengel) [13] developped a new algorithm to efficiently enumerates the solutions of certain type of circuits. They apply their result to give new proofs previous results on efficient enumeration for queries defined by tree automata or FO queries over structures with bounded tree width by using these circuits as aggregates to represent the set of all solutions of a query and then enumerating them.

Again at ***ICALP*** [15] Bacquey in an collaboration with Caen and Marseille (Grandjean and Olive) prove that linear time complexity on cellular automata is exactly characterized by inductive first-order Horn formulas. The method of proof also implies the following result: the enumeration of the ground atoms that are consequences of any inductive first-order Horn formula on a given structure can be performed in linear time (in the cardinality of the domain of the structure) by a cellular automaton (of appropriate dimension).

7.1.2. Provenance

Provenance is a type of aggregates that aims at exhibiting the contributions of tuples of a database to a query answer. This allows to give an explanation of the query answers, that can help to judge their reliability. Provenance is studied within the ANR project *Aggreg*.

In a paper at ICDT [14], Bourhis (with Amarilli, Monet and Senellart) studies the combined complexity for computing circuit representation of the provenance, which were used to efficiently evaluate aggregations tasks. In particular, they exhibit a recursive language of queries capturing path queries that compute a compact representation of the provenance.

7.1.3. Recursive Queries

At ***PODS*** [21], P. Bourhis proposed a formalisation of JSON documents, query languages and schema. This work is a collaboration with Chile. After having defined a clean theoretical framework to study JSON documents, Bourhis and his co-authors study the decidability and complexity of navigational query answering for different languages, relating each of them with existing implementations. Finally, they extend the documents with recursion together with a suitable querying language and study the complexity of query evaluating and query answering in this case.

At ***ICALP*** [17], P. Bourhis studied in a collaboration with Oxford the problem of definability in decidable fixpoint logic. Bourhis and his co-authors gives new characterisation of formulas that can be expressed in decidable logic with fixpoint. One of their main result is an effective characterisation of the formulas of the guarded negation fragment with fixpoint that can be expressed in the guarded fragment with fixpoint. Their techniques are then extended to effectively characterise the first order formulas that can be defined in the guarded fragment.

A. Lemay contributed at ICDE [16] the *gMark* benchmark, a tool to generate large size graph database and an associated set of queries. This work was done in cooperation with Eindhoven and previous members of Links that are now in Lyon and Clérmont-Ferrant. Its main interest is a great flexibility (the generation of the graph can be done from a simple schema, but can also incorporate elaborate a parameters), an ability to generate recursive queries, and the possibility to generate large sets of queries of a desired selectivity. This benchmark allowed for instance to highlight difficulties for the existing query engines to deal with recursive queries of high selectivity.

7.1.4. Data Integration

P. Bourhis and S. Tison presented at ***IJCAI*** [18] — the top conference in Artificial Intelligence — a new ontology mediated query answering system (OMQA) for JSON document. This work is a collaboration with researchers from the University of Montpellier. The strength of their contribution lies in the fact that their ontology is very expressive and yet gives a tractable query answering system. Moreover, they establish a non-trivial connection between their query answering system and term rewriting, allowing them to pinpoint the exact complexity of query answering and to evaluate it directly over KV-stores.

Also a ***IJCAI*** [20], P. Bourhis studied guarded ontology languages that are compatible with cross product. This work was done in cooperation with Edinburgh and Vienna. Cross product is a useful modelling tool that allow to connect every element of one relation to every element of another relation. However, in this paper, Bourhis and his co-authors show that its introduction into guarded ontology – even when it is limited to two relations – quickly leads to the undecidability of query evaluation and query answering. However, they isolate fragments where one can add cross products without losing the decidability of these problems by either restricting the queries or the ontology.

7.1.5. Schema Validation

I. Boneva presented at ISWC [19] her work on ShEx 2.0 (Shape Expression Language 2.0), a language to describe the vocabulary and the structure of an RDF graph. This work is a collaboration with Oviedo and MIT. The language is based on the notion of shapes, a typing system supporting algebraic operations, recursive references to other shapes or Boolean combination. In the paper, Boneva and her co-authors give efficient algorithms to test if an RDF graph satisfies a shapes schema together with implementation guidelines. Her research on the topic has also led to the publication of a book [25] on the validation of RDF data, containing among other things her contribution to ShEx.

JSON documents are basically unordered data trees. Schemas for unordered data trees can thus be defined by appropriate notions of tree automata for unordered trees, as studied in a systematic manner by Boiret, Hugot, and Niehren [11] in cooperation with Treinen from Paris 7. Alternatively, schemas can be defined by closed logic formulas in the logics proposed by the same authors in [12] .They showed that logics for unordered data trees with equality tests of data values of siblings nodes remain decidable, and thus the equivalence problems of the corresponding tree automata. In contrast, the problem becomes undecidable when comparing cousins for equality of data values.

7.2. Managing Dynamic Linked Data

7.2.1. Complex Event Processing

In his PhD project [24], Sakho supervised by Niehren and Boneva proposes studies the complexity of answering automata queries on hyperstreams. A hyperstream is collections of streams that are connected with each other. The motivation for hyperstreams is to avoid blocking when composing of several stream processes. They show that the problem of deciding whether a tuple can is certain for selection on a hyperstream by query defined by finite automaton is PSPACE-complete.

7.2.2. Transformations

Symbolic tree transducers define transformations of data trees Adrien Boiret, Vincent Hugot and Joachim Niehren could show at at DLT [23] that the equivalence problem of symbolic top-down tree transducers can be reduced to that of standard top-down tree transducers. Thereby, the existing equivalence testers can be lifted to the symbolic tree transducers, yielding the algorithms needed for verification tasks in the ANR projec CoLiS. An implementation of such an algorithm by Nicolas Bacquey is on the way.

P. Gallot and S. Salvati presented their work on 1-register streaming string transducers at ***STACS*** [22]. This work is a collaboration with University of Bordeaux. Streaming String Transducers have recently gained a growing interest since they can be used to model transformations on data streams. In this work, P. Gallot, S. Salvati and their co-authors prove that 1-register streaming string transducers can be decomposed as a finite union of functional transducers. An immediate corollary of this result is that the equivalence of such transducers is decidable, which means that we can check if two given transducers represent the same transformation on data streams.

MAGNET Project-Team

7. New Results

7.1. Natural Language Processing

In [13] we present a new, efficient method for learning task-specific word vectors using a variant of the Passive-Aggressive algorithm. Specifically, this algorithm learns a word embedding matrix in tandem with the classifier parameters in an online fashion, solving a bi-convex constrained optimization at each iteration. We provide a theoretical analysis of this new algorithm in terms of regret bounds, and evaluate it on both synthetic data and NLP classification problems, including text classification and sentiment analysis. In the latter case, we compare various pre-trained word vectors to initialize our word embedding matrix, and show that the matrix learned by our algorithm vastly outperforms the initial matrix, with performance results comparable or above the state-of-the-art on these tasks.

In [12] we present a new approach to the problem of cross-lingual dependency parsing, aiming at leveraging training data from different source languages to learn a parser in a target language. Specifically, this approach first constructs word vector representations that exploit structural (i.e., dependency-based) contexts but only considering the morpho-syntactic information associated with each word and its contexts. These delexicalized word em-beddings, which can be trained on any set of languages and capture features shared across languages, are then used in combination with standard language-specific features to train a lexicalized parser in the target language. We evaluate our approach through experiments on a set of eight different languages that are part the Universal Dependencies Project. Our main results show that using such delexicalized embeddings, either trained in a monolingual or multilingual fashion, achieves significant improvements over monolingual baselines.

7.2. Decentralized Learning and Privacy

In [15] we consider a set of learning agents in a collaborative peer-to-peer network, where each agent learns a personalized model according to its own learning objective. The question addressed in this paper is: how can agents improve upon their locally trained model by communicating with other agents that have similar objectives? We introduce and analyze two asynchronous gossip algorithms running in a fully decentralized manner. Our first approach, inspired from label propagation, aims to smooth pre-trained local models over the network while accounting for the confidence that each agent has in its initial model. In our second approach, agents jointly learn and propagate their model by making iterative updates based on both their local dataset and the behavior of their neighbors. To optimize this challenging objective, our decentralized algorithm is based on ADMM.

In subsequent work in collaboration with Rachid Guerraoui's group at EPFL [18], we study how agents can collaborate to improve upon their locally learned model without leaking sensitive information about their data. Our first contribution is to reformulate this problem so that it can be solved by a block coordinate descent algorithm. We obtain an efficient and fully decentralized protocol working in an asynchronous fashion. Our second contribution is to make our algorithm differentially private to protect against the disclosure of any information about personal datasets. We prove convergence rates and exhibit the trade-off between utility and privacy. Our experiments show that our approach dramatically outperforms previous work in the non-private case, and that under privacy constraints we significantly improve over purely local models. A preliminary version of this work was presented at the NIPS 2017 workshop on machine Learning on the Phone and other Consumer Devices [16].

7.3. Statistical Learning on Graphs

The main purpose of [11] is to illustrate that certain Hölder-type inequalities can be employed in order to obtain concentration and correlation bounds for sums of weakly dependent random variables whose dependencies are described in terms of graphs, or hypergraphs. Let Y_v , $v \in V$, be real-valued random variables having a dependency graph $G = (V, E)$. We show that

$$\mathbb{E}\left[\prod_{v\in V} Y_v\right] \leq \prod \left\{\mathbb{E}\left[Y_v^{\frac{\chi_b}{b}}\right]\right\}^{\frac{b}{\chi_b}}$$

where χ_b is the b-fold chromatic number of G. This inequality may be seen as a dependency-graph analogue of a generalized Hölder inequality, due to Helmut Finner. Additionally, we provide applications of the aforementioned Hölder-type inequalities to concentration and correlation bounds for sums of weakly dependent random variables whose dependencies can be described in terms of graphs or hypergraphs.

Several collaborations concerned efficient counting of subgraph frequencies in networks. Two journal articles are accepted subject to minor revisions, one in collaboration with the group of Yvan Saeys (University of Ghent, Belgium), and one in collaboration with Irma Ravkic and Martin Znidarsic (former collaborators of JAN RAMON).

7.4. Data Mining with Rank Data

Rank data, in which each row is a complete or partial ranking of available items (columns), is ubiquitous. Among others, it can be used to represent preferences of users, levels of gene expression, and outcomes of sports events. It can have many types of patterns, among which consistent rankings of a subset of the items in multiple rows, and multiple rows that rank the same subset of the items highly. In [10], we show that the problems of finding such patterns can be formulated within a single generic framework that is based on the concept of semiring matrix factorization. In this framework, we employ the max-product semiring rather than the plus-product semiring common in traditional linear algebra. We apply this semiring matrix factorization framework on two tasks: sparse rank matrix factorization and rank matrix tiling. Experiments on both synthetic and real world datasets show that the framework is capable of discovering different types of structure as well as obtaining high quality solutions.

7.5. Large-Scale Machine Learning

In [19], we study large-scale kernel methods for acoustic modeling in speech recognition and compare their performance to deep neural networks (DNNs). We perform experiments on four speech recognition datasets and compare these two types of models on frame-level performance metrics (accuracy, cross-entropy), as well as on recognition metrics (word/character error rate). In order to scale kernel methods to these large datasets, we use the random Fourier feature method. We propose two novel techniques for improving the performance of kernel acoustic models. First, in order to reduce the number of random features required by kernel models, we propose a simple but effective method for feature selection. Second, we present a number of frame-level metrics which correlate very strongly with recognition performance when computed on the heldout set; we take advantage of these correlations by monitoring these metrics during training in order to decide when to stop learning. Additionally, we show that the linear bottleneck method of Sainath et al. improves the performance of our kernel models significantly, in addition to speeding up training and making the models more compact. Together, these three methods dramatically improve the performance of kernel acoustic models, making their performance comparable to DNNs on the tasks we explored.

7.6. Beyond Homophily: Signed networks

In the problem of edge sign prediction, we are given a directed graph (representing a social network), and our task is to predict the binary labels of the edges (i.e., the positive or negative nature of the social relationships). Many successful heuristics for this problem are based on the troll-trust features, estimating at each node the fraction of outgoing and incoming positive/negative edges. In [14], we show that these heuristics can be understood, and rigorously analyzed, as approximators to the Bayes optimal classifier for a simple probabilistic model of the edge labels. We then show that the maximum likelihood estimator for this model approximately corresponds to the predictions of a Label Propagation algorithm run on a transformed version of the original social graph. Extensive experiments on a number of real-world datasets show that this algorithm is competitive against state-of-the-art classifiers in terms of both accuracy and scalability. Finally, we show that troll-trust features can also be used to derive online learning algorithms which have theoretical guarantees even when edges are adversarially labeled.

MINT2 Team

6. New Results

6.1. Localized Haptic Texture: A Rendering Technique Based on Taxels for High Density Tactile Feedback

Participants: Yosra Rekik, Eric Vezzoli, Laurent Grisoni, Frederic Giraud.

We investigate the relevance of surface haptic rendering techniques for tactile devices. We focus on the two major existing techniques and show that they have complementary benefits. The first one, called Surface Haptic Object (SHO), which is based on finger position, is shown to be more suitable to render sparse textures; while the second one, called Surface Haptic Texture (SHT), which is based on finger velocity, is shown to be more suitable for dense textures and fast finger movements. We hence propose a new rendering technique, called Localized Haptic Texture (LHT), which is based on the concept of *taxel* considered as an elementary tactile information that is rendered on the screen. By using a grid of taxels to encode a texture, LHT is shown to provide a consistent tactile rendering across different velocities for high density textures, and is found to reduce user error rate by up to 77.68% compared to SHO. Available at https://hal.inria.fr/hal-01444099.

6.2. Revgest: Augmenting Gestural Musical Instruments with Revealed Virtual Objects

Participants: Florent Berthaut, Cagan Arslan, Laurent Grisoni.

Gestural interfaces, which make use of physiological signals, hand / body postures or movements, have become widespread for musical expression. While they may increase the transparency and expressiveness of instruments, they may also result in limited agency, for musicians as well as for spectators. This problem becomes especially true when the implemented mappings between gesture and music are subtle or complex. These instruments may also restrict the appropriation possibilities of controls, by comparison to physical interfaces. Most existing solutions to these issues are based on distant and/or limited visual feedback (LEDs, small screens). Our approach is to augment the gestures themselves with revealed virtual objects. Our contributions are, first a novel approach of visual feedback that allow for additional expressiveness, second a software pipeline for pixel-level feedback and control that ensures tight coupling between sound and visuals, and third, a design space for extending gestural control using revealed interfaces. We also demonstrate and evaluate our approach with the augmentation of three existing gestural musical instruments. Available at URL https://hal.archives-ouvertes.fr/hal-01518579

6.3. Spontaneous Gesture Production Patterns on Multi-Touch Interactive Surfaces

Participants: Yosra Rekik, Radu-Daniel Vatavu, Laurent Grisoni.

Expressivity of hand movements is much greater than what current interaction techniques enable in touch-screen input. Especially for collaboration, hands are used to interact but also to express intentions, point to the physical space in which collaboration takes place, and communicate meaningful actions to collaborators. Various types of interaction are enabled by multi-touch surfaces (singe and both hands, single and multiple fingers, etc), and standard approaches to tactile interactive systems usually fail in handling such complexity of expression. The diversity of multi-touch input also makes designing multi-touch gestures a difficult task. We believe that one cause for this design challenge is our limited understanding of variability in multi-touch gesture articulation, which affects users' opportunities to use gestures effectively in current multi-touch interfaces. A better understanding of multi-touch gesture variability can also lead to more robust design to support different users' gesture preferences. In this work we present our results on multi-touch gesture

variability. We are mainly concerned with understanding variability in multi-touch gestures articulation from a pure user-centric perspective. We present a comprehensive investigation on how users vary their gestures in multi-touch gestures even under unconstrained articulation conditions. We conducted two experiments from which we collected 6669 multi-touch gestures from 46 participants. We performed a qualitative analysis of user gesture variability to derive a taxonomy for users' multi-touch gestures that complements other existing taxonomies. We also provide a comprehensive analysis on the strategies employed by users to create different gesture articulation variations for the same gesture type. Available at URL https://hal.inria.fr/hal-01444113

6.4. Control and evaluation of a 2-D Multimodal Controlled-Friction Display

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

The multimodal control of a 2D controlled-Friction Device is presented. We use the Vector control method because the phase and amplitude of two vibration modes at a same frequency can be precisely set. The closed loop response time of 10 ms is achieved. The device is then associated with a finger position sensor. The algorithm of the multimodal approach is then tested. In spite of the inevitable limitations of the system-saturation of amplifiers, low sampling frequency of the sensor-low friction could be imposed under a finger while a high friction was imposed on a predetermined position. This confirms the validity of the modal approach to create multi touch interactions. Available at URL https://hal.inria.fr/hal-01538340

6.5. Enriching Musical Interaction on Tactile Feedback Surfaces with Programmable Friction

Participants: Farzan Kalantari, Florent Berthaut, Laurent Grisoni.

In the recent years, a great interest has emerged to utilise tactile interfaces for musical interactions. These interfaces can be enhanced with tactile feedback on the user's fingertip through various technologies, including programmable friction techniques. In this study, we have used a qualitative approach to investigate the potential influence of these tactile feedback interfaces on user's musical interaction. We have experimented three different mappings between the sound parameters and the tactile feedback in order to study the users' experiences of a given task. Our preliminary findings suggest that friction-based tactile feedback is a useful tool to enrich musical interactions and learning. Available at URL https://hal.inria.fr/hal-01580750

6.6. bf-pd: Enabling Mediated Communication and Cooperation in Improvised Digital Orchestras

Participants: Luke Dahl, Florent Berthaut, Antoine Nau, Patricia Plenacoste.

Digital musical instruments enable new musical collaboration possibilities, extending those of acoustic ensembles. However, the use of these new possibilities remains constrained due to a lack of a common terminology and technical framework for implementing them. Bf-pd is a new software library built in the PureData (Pd) language which enables communication and cooperation between digital instruments. Its design is based on the BOEUF conceptual framework which consists of a classification of modes of collaboration used in collective music performance, and a set of components which affords them. Bf-pd can be integrated into any digital instrument built in Pd, and provides a "collaboration window" from which musicians can easily view each others' activity and share control of instrument parameters and other musical data. We evaluate the implementation and design of bf-pd through workshops and a preliminary study and discuss its impact on collaboration within improvised ensembles of digital instruments. Available at URL https://hal.archives-ouvertes.fr/hal-01577942

6.7. Toward Augmented Familiarity of the Audience with Digital Musical Instruments

Participants: Olivier Capra, Florent Berthaut, Laurent Grisoni.

The diversity and complexity of Digital Musical Instruments often lead to a reduced appreciation of live performances by the audience. This can be linked to the lack of familiarity they have with the instruments. We propose to increase this familiarity thanks to a trans-disciplinary approach in which signals from both the musician and the audience are extracted, familiarity analysed, and augmentations dynamically added to the instruments. We introduce a new decomposition of familiarity and the concept of correspondences between musical gestures and results. This study is both a review of research that paves the way for the realisation of a pipeline for augmented familiarity, and a call for future research on the identified challenges that remain before it can be implemented. Available at URL https://hal.archives-ouvertes.fr/hal-01577953

6.8. Understanding Gesture Articulations Variability

Participants: Orlando Erazo, Yosra Rekik, Laurent Grisoni, José Pino.

Interfaces based on mid-air gestures often use a one-to-one mapping between gestures and commands, but most remain very basic. Actually, people exhibit inherent intrinsic variations for their gesture articulations because gestures carry dependency with both the person producing them and the specific context, social or cultural, in which they are being produced. We advocate that allowing applications to map many gestures to one command is a key step to give more flexibility, avoid penalisations, and lead to better user interaction experiences. Accordingly, this study presents our results on mid-air gesture variability. We are mainly concerned with understanding variability in mid-air gesture articulations from a pure user-centric perspective. We describe a comprehensive investigation on how users vary the production of gestures under un-constrained articulation conditions. The conducted user study consisted in two tasks. The first one provides a model of user conception and production of gestures; from this study we also derive an embodied taxonomy of gestures. This taxonomy is used as a basis for the second experiment, in which we perform a fine grain quantitative analysis of gesture articulation variability. Based on these results, we discuss implications for gesture interface designs. Available at URL https://hal.inria.fr/hal-01578738

6.9. Understanding Users' Perception of Simultaneous Tactile Textures

Participants: Yosra Rekik, Eric Vezzoli, Laurent Grisoni.

We study users' perception of simultaneous tactile textures in ultrasonic devices. We investigate how relevant is providing the user with different complementary and simultaneous textures with respect to the different fingers that can be used to touch the surface. We show through a controlled experiment that users are able to distinguish the number of different textures independently of using fingers from one or two hands. However, our findings indicate that users are not able to differentiate between two different textures, that is to correctly identify each of them, when using fingers from the same hand. Based on our findings, we are then able to outline three relevant guidelines to assist multi-finger tactile feedback ergonomic and devices design. Available at URL https://hal.inria.fr/hal-01578729

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. Understanding and modeling users

Participants: Géry Casiez [correspondent], Stéphane Huot, Edward Lank, Justin Dan Mathew, Mathieu Nancel, Sylvain Malacria, Nicolas Roussel, Marcelo Wanderley.

7.2.1. Understanding the practices of 3D audio production professionals

3D audio production tools vary from low-level programming libraries to higher-level user interfaces that are used across a wide range of applications. However, many of the user interfaces for authoring 3D audio parameters are underdeveloped, forcing users to resort to ad hoc solutions with other tools or programming languages, which limits creativity and productivity. Even though there is a significant increase of interest in this problem, usability issues with the manipulation of spatial parameters in 3D audio tools are still not well identified. We have thus conducted an on-line survey with practitioners to gather ethnographic information on their tools, methods, and assessments [15]. Our goal was to identify limitations and custom methods to circumvent them, in order to inform the development of better user interfaces for 3D audio production. Results of the survey revealed specific methods and limitations with regards to *Audio Rendering & Recording*, *Visual Feedback*, *Functionality*, and *Workflow Integration*. We also identified three basic but important tasks that have to be performed interactively with 3D audio production tools: *Defining the Rendering Space*, *Creation and Manipulation of Audio Objects*, and *Monitoring with Audio/Visual Feedback*. As part of Justin Mathew's PhD [11], this classification helped to identify the needs and to experiment new 3D audio tools that address issues with low-level functionality of 3D audio production systems such as the specification of 3D audio trajectories.

7.2.2. Exploring playfulness in the design of new musical instruments

Play and playfulness compose an essential part of our lives as human beings. From childhood to adulthood, playfulness is often associated with remarkable positive experiences related to fun, pleasure, intimate social activities, imagination, and creativity. Not surprisingly, playfulness has been recurrently considered an important criterion in the design of interfaces for Digital Musical Instruments. It is supposed to engage people, often non-expert, in short term musical activities. Yet, designing for playfulness remains a challenging task, as little knowledge is available for designers to support their decisions. To address this issue, we have followed a design rationale approach using the context of Live Looping as a case study [20]. We first surveyed 101 Live Looping tools and summarized our analysis into a new design space. We used this design space to discuss potential guidelines to address playfulness in a design process: (i) advanced looping capacity, so as to extend the musical possibilities of Live Looping; (ii) low input capacity and direct mappings, in order to help getting familiar with the instrument and to favor direct control; and (iii) transparent and intense visual feedback to help infer what is happening inside the device when an action is performed. These guidelines were implemented and discussed in a new prototype of Live Looping instrument, the *Voice Reaping Machine*.

7.2.3. Use of brain-computer interfaces

BCIs are presumably supposed to require the full attention of their users, and to lose accuracy if users pay attention to another task. This assertion has been verified with several BCI paradigms (e.g. P300). But the cognitive demand of the promising SSVEP paradigm had never been specifically assessed. We measured the accuracy of an SSVEP-based BCI used by 26 participants in various conditions of mental workload [22]. Our analysis revealed that surprisingly, for this type of BCI, little attention is actually needed from participants to reach optimal accuracy: participants were able to successfully perform a complex secondary task (N-back) without degrading the BCI accuracy. The same observation was made whether visual or auditive attention was solicited. These results indicate that SSVEP is a low-demanding paradigm in terms of cognitive resources, and are encouraging for its use in complex interaction settings. Last we did a survey on the use of BCIs in augmented reality. The results of the survey show that most of the previous work made use of P300 or SSVEP paradigms with EEG in Video See-Through systems, and that robotics is a main field of application with the highest number of existing systems [33].

7.2.4. Modeling user performance on curved constrained paths

In 2D and 3D interfaces, a "steering task" consists in following a predefined path of arbitrary shape, with a given and possibly varying tolerance (or "width"), using the finger or the cursor. While less common than target acquisition, this family of pointing tasks is a distinct component of modern interaction, e.g. selecting items in a hierarchical linear menus, or deep-etching an image in Photoshop. Previous work have essentially modeled straight or circular paths of constant width, and argued that more complex paths can be modeled using combinations of these primitives. We demonstrated that existing models actually fail to correctly model constrained paths of varying, arbitrary curvature [29]. We proposed a new model that integrates instantaneous curvature and width into its formulation, and validated it empirically for direct touch.

7.3. Interactive visualization and animations

Participants: Amira Chalbi-Neffati, Fanny Chevalier [correspondent], Nicolas Roussel.

7.3.1. Visualization literacy at elementary school

This work advances our understanding of children's visualization literacy, and aims to improve it with a novel approach for teaching visualization at elementary schools. We contributed an analysis of data graphics and activities employed in grade K to 4 educational materials, and conducted a survey with 16 elementary school teachers. We found that visualization education could benefit from integrating pedagogical strategies for teaching abstract concepts with established interactive visualization techniques. Building on these insights, we have developed and studied design principles for novel interactive teaching material aimed at increasing children's visualization literacy. We developed *C'est la vis*, an online platform for teachers and students to respectively teach and learn about pictographs and bar charts (see Fig. 1), and reported on our initial observations of its use in grades K and 2 [16]. The application can be tested at https://cestlavis.github.io. This work was awarded a Best Paper Honorable Mention award at the ACM CHI conference and highlighted on FlowingData and Data Stories, two famous blogs on visualization.

7.3.2. Supporting handoff in asynchronous collaborative sensemaking

During asynchronous collaborative analysis, handoff of partial findings is challenging because externalizations produced by analysts may not adequately communicate their investigative process. To address this challenge, we developed techniques to automatically capture and help encode tacit aspects of the investigative process based on an analyst's interactions, and streamline explicit authoring of handoff annotations []. We designed our techniques to mediate awareness of analysis coverage, support explicit communication of progress and uncertainty with annotation, and implicit communication through playback of investigation histories. To evaluate our techniques, we developed an interactive visual analysis system, *KTGraph*, that supports an asynchronous investigative document analysis task. We conducted a two-phase user study to characterize a set of handoff strategies and to compare investigative performance with and without our techniques. The results suggest that our techniques promote the use of more effective handoff strategies, help increase an awareness

Figure 1. C'est la Vis is a tablet-based application co-designed with elementary teachers to support the teaching and learning of pictographs and bar charts in Grades K-4.

of prior investigative process and insights, as well as improve final investigative outcomes. This work received a Best Paper Honorable Mention award at the IEEE VIS-VAST conference.

7.3.3. Phenotype comparison visualizations for disease subtyping via topic models

PhenoLines is a visual analysis tool for the interpretation of disease subtypes, derived from the application of topic models to clinical data [25]. Topic models enable one to mine cross-sectional patient comorbidity data (e.g., electronic health records) and construct disease subtypes—each with its own temporally evolving prevalence and co-occurrence of phenotypes—without requiring aligned longitudinal phenotype data for all patients. However, the dimensionality of topic models makes interpretation challenging, and *de facto* analyses provide little intuition regarding phenotype relevance or phenotype interrelationships. PhenoLines allows to compare phenotype prevalence within and across disease subtype topics (see Fig. 2), thus supporting subtype characterization, a task that involves identifying a proposed subtype's dominant phenotypes, ages of effect, and clinical validity. We contributed a data transformation workflow that employs the Human Phenotype Ontology to hierarchically organize phenotypes and aggregate the evolving probabilities produced by topic models. We introduced a novel measure of phenotype relevance that can be used to simplify the resulting topology. The design of PhenoLines was motivated by formative interviews with machine learning and clinical experts. We conducted initial evaluations with machine learning experts and a medical domain expert. Results suggest that PhenoLines supports promising approaches for the characterization and optimization of topic models.

7.3.4. Understanding and designing animation for visualization

Animations are increasingly used in interactive systems in order to enhance the usability and aesthetics of user interfaces. While animations are proven to be useful in many cases, we still observe defective ones causing many problems, such as distracting users from their main task or making data exploration slower. The fact that such animations still exist proves that animations are not yet very well understood as a cognitive aid, and that we have not yet definitely decided what makes a well designed one. Our work on this topic aims at better understanding the different aspects of animations for user interfaces and exploring new methods and guidelines for designing them.

We explored the concept of *dataTours*, semi-automated narratives for prompting and sustaining exploratory analysis, as a complement to the predominant interactive data overview. Narratives are commonly and widely used for presenting and communicating results of data analyses and thus generally lead an observer to a specific resolution or conclusion regarding the question(s) they address. A dataTour, on the other hand, is meant to be an open-ended and adaptive narrative that can act as a scaffolding for the iterative process of exploration. Preliminary results were presented as a poster at the IEEE Infovis conference [37].

7.4. Interaction techniques

Figure 2. PhenoLines facilitates the visual analysis of topics that describe disease symptoms, in support of topic model optimization and characterization. Hierarchical relationships, temporal trends, correlated measures, and rank-ordered lists enable for comparisons within and between topics.

Participants: Axel Antoine, Géry Casiez, Fanny Chevalier, Stéphane Huot, Sylvain Malacria [correspondent], Thomas Pietrzak, Sébastien Poulmane, Nicolas Roussel.

7.4.1. Touch interaction with finger identification: which finger(s) for what?

HCI researchers lack low latency and robust systems to support the design and development of interaction techniques that leverage finger identification. We developed a low cost prototype, called *WhichFingers* (see Fig. 3), using piezo based vibration sensors attached to each finger [28]. By combining the events from an input device with the information from the vibration sensors we demonstrated how to achieve low latency and robust finger identification. We evaluated our prototype in a controlled experiment, showing recognition rates of 98.2% for keyboard typing, and 99.7% for single touch and 94.7% for two simultaneous touches on a touchpad. These results were confirmed in an additional laboratory experiment with ecologically valid tasks.

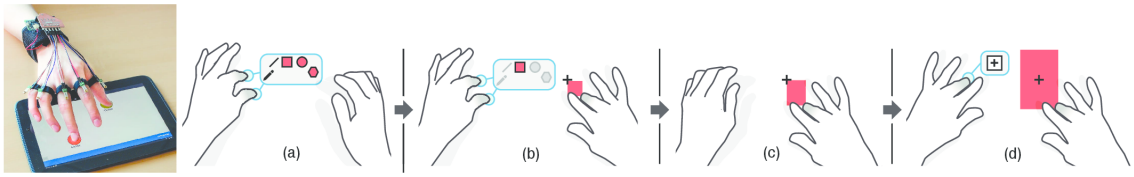

Figure 3. (left) WhichFingers wireless device prototype. (right) Example of how FingerCuts merges command selection with direct manipulation of command parameters with feed-forward: (a) specific finger chords with non-dominant hand displays possible commands; (b) dominant hand selects command using specific corresponding finger, e.g. middle finger triggers "start drawing rectangle"; (c) dominant hand movement provides continuous control of command parameters, e.g. size of the rectangle; (d) other specific fingers with non-dominant hand further tunes the command parameters, e.g. middle finger change mode to "draw rectangle from the center".

We also explored the large input space made possible by such finger identification technology. *FingerCuts* for instance is an interaction technique inspired by desktop keyboard shortcuts [14]. It enables integrated command selection and parameter manipulation, it uses feed-forward and feedback to increase discoverability, it is backward-compatible with current touch input techniques, and it is adaptable for different form factors of

touch devices: tabletop, tablet, and smartphone. Qualitative and quantitative studies conducted on a tabletop suggest that with some practice, FingerCuts is expressive, easy to use, and increases a sense of continuous interaction flow. Interaction with FingerCuts was found as fast or faster than using a graphical user interface. A theoretical analysis of FingerCuts using the Fingerstroke-Level Model (FLM) [42] matches our quantitative study results, justifying our use of FLM to analyze and estimate the performance for other device form factors.

7.4.2. Force-based autoscroll

Autoscroll, also known as edge-scrolling, is a common interaction technique in graphical interfaces that allows users to scroll a viewport while in dragging mode: once in dragging mode, the user moves the pointer near the viewport's edge to trigger an "automatic" scrolling. In spite of their wide use, existing autoscroll methods suffer from several limitations [38]. First, most autoscroll methods over-rely on the size of the control area, that is, the larger it is, the faster scrolling rate can be. Therefore, the level of control depends on the available distance between the viewport and the edge of the display, which can be limited. This is for example the case with small displays or when the view is maximized. Second, depending on the task, the users' intention can be ambiguous (e.g. dragging and dropping a file is ambiguous as the user's target may be located within the initial viewport or in a different one on the same display). To reduce this ambiguity, the size of the control area is drastically smaller for drag-and-drop operations which consequently also affects scrolling rate control as the user has a limited input area to control the scrolling speed.

We presented a conceptual framework of factors influencing the design of edge-scrolling techniques [12]. We then analyzed 33 different desktop implementations by reverse-engineering their behavior, and demonstrated substantial variance in their design approaches. Results of an interactive survey with 214 participants show that edge-scrolling is widely used and valued, but also that users encounter problems with control and with behavioral inconsistencies. Finally, we reported results of a controlled experiment comparing four different implementations of edge-scrolling, which highlight factors from the design space that contribute to substantial differences in performance, overshooting, and perceived workload.

We also explored how force-sensing input, which is now available on several commercial devices, can be used to overcome the limitations of autoscroll. Indeed, force-sensing is an interesting candidate because: 1) users are often already applying a (relatively soft) force on the input device when using autoscroll and 2) varying force on the input device does not require to move the pointer, thus making it possible to offer control to the user while using a small and consistent control area regardless of the task and the device. We designed *ForceEdge*, a novel interaction technique mapping the force applied on a trackpad to the autoscrolling rate [17]. We implemented a software interface that can be used to design different transfer functions that map the force to autoscrolling rate and test these mappings for text selection and drag-and-drop tasks. The results of three controlled experiments suggest that it improves over MacOS and iOS systems baselines for top-to-bottom select and move tasks.

7.4.3. Free-space gestural interaction

7.4.3.1. Exploring at-your-side gestural interaction for ubiquitous environments

Free-space or in-air gestural systems are faced with two major issues: a lack of subtlety due to explicit mid-air arm movements, and the highly effortful nature of such interactions. The lack of subtlety can influence whether gestures are used; specifically, if gestures require large-scale movement, social embarrassment can restrict their use to private contexts. Similarly, large-scale movements are tiring, further limiting gestural input to short-term tasks by physically health users.

We believe that, to promote gestural input, lower-effort and more socially acceptable interaction paradigms are essential. To address this need, we explored at-one's-side gestural input, where the user gestures with their arm down at their side in order to issue commands to external displays. Within this space, we presented the results of two studies that investigate the use of side-gesture input for interaction [34]. First, we investigated end-user preference through a gesture elicitation study, presented a gesture set, and validated the need for dynamic, diverse, and variable-length gestures. We then explored the feasibility of designing such a gesture recognition system, dubbed *WatchTrace*, which supports alphanumeric gestures of up to length three with an average accuracy of up to 82%, providing a rich, dynamic, and feasible gestural vocabulary.

7.4.3.2. Effect of motion-gesture recognizer error pattern on user workload and behavior

In free-space gesture recognition, the system receiving gesture input needs to interpret input as the correct gesture. We measure a system's ability to do this using two measures: precision, or the number of gestures classified correctly versus the number of gestures misclassified; and recall, or the overall number of gestures of a given class that are classified. To promote precision, systems frequently require gestures to be performed more accurately, and systems then reject less careful gesture as input, forcing the user to retry the action. To promote this accuracy, the system sets a threshold, a criterion value, that describes how accurately the gesture must be performed. If we tighten this threshold, we increase the precision of the system, but this, in turn, results in more rejected gestures, negatively impacting recall.

Bi-level thresholding is a motion gesture recognition technique that mediates between precision and recall using two threshold levels: a tighter threshold that limits recognition errors or boost precision, and a looser threshold that promotes higher recall. These two thresholds accomplish this goal by analyzing movements in sequence and treating two less-precise gestures the same as a single more precise gesture, i.e. two near-miss gestures is interpreted the same as a single accurately performed gesture. We explored the effects of bi-level thresholding on the workload and acceptance of end-users [27]. Using a wizard-of-Oz recognizer, we held recognition rates constant and adjusted for fixed versus bi-level thresholding. Given identical recognition rates, we showed that systems using bi-level thresholding result in significant lower workload scores (on the NASA-TLX) and significantly lower accelerometer variance.

7.4.3.3. Rapid interaction with interface controls in mid-air

Freehand interactions with large displays often rely on a "point & select" paradigm. However, constant hand movement in air for pointer navigation quickly leads to arm and hand fatigue. We introduced *summon & select*, a new model for freehand interaction where, instead of navigating to the control, the user summons it into focus and then manipulates it [26]. Summon & select solves the problems of constant pointer navigation, need for precise selection, and out-of-bounds gestures that plague point & select. We conducted two studies to evaluate the design and compare it against point & select in a multi-button selection study. Results suggest that summon & select helps performing faster and has less physical and mental demand than point & select.

7.4.4. Using toolbar button icons to communicate keyboard shortcuts

Toolbar buttons are frequently-used widgets for selecting commands. They are designed to occupy little screen real estate, yet they convey a lot of information to users: the icon is directly tied to the meaning of the command, the color of the button informs whether the command is available or not, and the overall shape and shadow effect together afford a point & click interaction to execute the command. Most commands attached to toolbars can also be selected by using an associated keyboard shortcut. Keyboard shortcuts enable users to reach higher performance than selecting a command through pointing and clicking, especially for frequent actions such as repeated "Copy/Paste" operations. However, keyboard shortcuts suffer from a poor accessibility. We proposed a novel perspective on the design of toolbar buttons that aims to increase keyboard shortcut accessibility [24]. *IconHK* implements this perspective by blending visual cues that convey keyboard shortcut information into toolbar buttons without denaturing the pictorial representation of their command (see Fig. 4). We introduced three design strategies to embed the hotkey, a visual encoding to convey the modifiers, and a magnification factor that determines the blending ratio between the pictogram of the button and the visual representation of the hotkey. Two studies explored the benefits of IconHK for end-users and provided insights from professional designers on its practicality for creating iconsets. Based on these insights, we built a tool to assist designers in applying the IconHK design principles.

7.5. Novel interactive technologies

Participants: Géry Casiez, Fanny Chevalier, Stéphane Huot [correspondent], Sylvain Malacria, Thomas Pietrzak, Sébastien Poulmane, Thibault Raffaillac, Nicolas Roussel, Marcelo Wanderley.

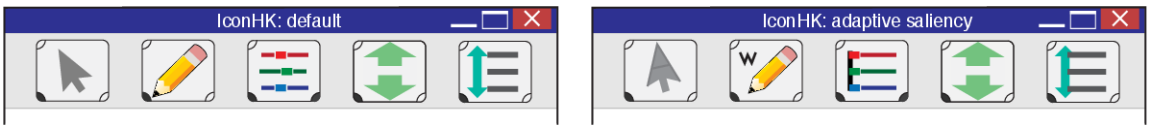

Figure 4. A toolbar using IconHK. (left) By default, the magnification of all buttons is at the default level and the hotkey symbol is not emphasized. (right) The magnification level of each button reflects user's command selection habits: the more frequently a command is selected by clicking, the more the hotkey symbol is emphasized.

7.5.1. Technologies for prototyping, tuning and studying interactive systems

7.5.1.1. Measure, characterization and compensation of end-to-end latency

We developed a low-cost method to measure and characterize the end-to-end latency when using a touch system (tap latency) or an input device equipped with a physical button [21]. Our method relies on a vibration sensor attached to a finger and a photo-diode to detect the screen response (see Fig. 5). Both are connected to a micro-controller connected to a host computer using a low-latency USB communication protocol in order to combine software and hardware probes to help determine where the latency comes from. We presented the operating principle of our method and investigated the main sources of latency in several systems. We showed that most of the latency originates from the display side. Our method can help application designers characterize and troubleshoot latency on a wide range of interactive systems.

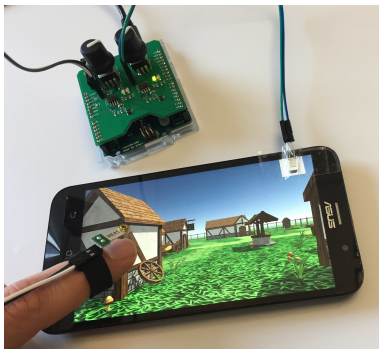

Figure 5. The hardware of our latency measuring method comprises a vibration sensor mounted on the finger to measure when a surface is touched or when a physical button is pressed, and a photo-diode to determine when the screen is updated.

When there is no other way to reduce latency (e.g. hardware improvements or software optimization), we have to rely on latency compensation methods. In direct interaction, this problem is considered and formulated as a trajectory prediction problem, i.e. where should the active position be without latency? We addressed this problem by constructing a frequency-domain approximation of the non-casual ideal predictor [18]. This approximation can be computed analytically, or obtained as an optimization task. We also proposed an adaptive modification of the algorithm taking into account possible variations in user behavior, and we empirically illustrated its applicability.

7.5.1.2. Turning function calls into animations

Animated transitions are an integral part of modern interaction frameworks. While early systems would animate a few properties, such as position or color, with different functions for each, modern systems contain too many properties to scale this way. To cope with this increasing numbers of properties to animate, most frameworks rely on naming strategies to refer to properties – like "position" or "color" – instead of having one specific function for each. This improves flexibility for choosing animated properties at runtime and reduces API size. It also gives an implicit contract that any property can animate. This flexibility has a price though. The animation of custom properties and types requires frameworks to provide an advanced API that exposes low-level details of their animation systems (e.g., timers and threads). This results in larger APIs and cumbersome syntaxes owing to the complex underlying mechanisms. It also creates a steep learning curve from basic to advanced APIs, which is likely to force programmers to stick to existing animatable properties whenever possible.

We have studied better ways for animating all objects in the system, independently of any frameworks, and aimed for a syntax reusing as much as possible the existing lexical elements [32]. This resulted in a delay operator appended to function calls – `object.setProperty(target)` *`during 2s`* – that turns setters into animations. It offers a coherent way to express animations across frameworks, and facilitates the animation of new properties. Our proof-of-concept implementation has been done in the Smalltalk language and tested with three popular interaction frameworks for the Pharo platform. In the future, this prototype should contribute to measuring the effect of animation syntax on programmers' ease in prototyping.

7.5.1.3. Interactive tuning of parameters without code recompilation

We developed *libParamTuner*, a cross-platform library that allows interactive tuning of parameters in applications written in C++ or Java, without the need to re-compile [19]. libParamTuner provides a lightweight syntax to bind some variables of an application to parameters defined in an XML file. Each modification of the XML file updates in real time the associated parameters in the application. A graphical interface allows editing the XML file, using interactive controls dynamically created for each parameter.

7.5.2. Technologies for creative applications

7.5.2.1. Leveraging flexion input for pen interaction

For artists, digital media offers some advantages over physical media, such as non-destructive editing, automation, and potential novel effects. However, conveying artistic intent to a computer system can be problematic. An artist working in traditional media, such as a painter or pastellist, can subtly manipulate their artistic tool to create different effects, while a digital artist working with a non-augmented stylus is only able to indicate a path, shape, or point on the surface of the tablet. Augmented styluses help solve this problem by using physical inputs such as pressure or tilt to function as parametric controls. We designed and developed *FlexStylus*, a flexible stylus that detects deformation of the barrel as a vector with both a rotational and an absolute value [23]. It provides two additional degrees of freedom with the goal of improving the expressiveness of digital art using a stylus device. The prototype uses a cluster of four fiber-optic-based deformation sensors. We also proposed interaction techniques using the FlexStylus to improve menu navigation and tool selection (see Fig. 6). Finally, we conducted a study comparing users' ability to match a changing target value using a commercial pressure stylus and the FlexStylus' absolute deformation. When using the FlexStylus, users had higher accuracy overall, suggesting that deformation may be a useful input method for future work considering stylus augmentation.

7.5.2.2. Versioning and annotation support for collaborative mapping design

A crucial component of the digital musical instrument (DMI) building process involves the mapping of sensor or gesture input signals from the musician to relevant synthesis parameters. While there are many different approaches to the design and implementation of DMI mapping, for many designers, a common part of the process involves the use of graphical user interfaces where the connection of signals can be observed and manipulated visually. In that context, and especially during collaborative design, being able to effectively store, preview, and reload previous configurations can reduce risks, encourage exploration, and improve the management of configuration files. We developed a new mapping tool for the libmapper library, a well-used system for connecting signals employed for the design and implementation of DMIs, that deeply integrate

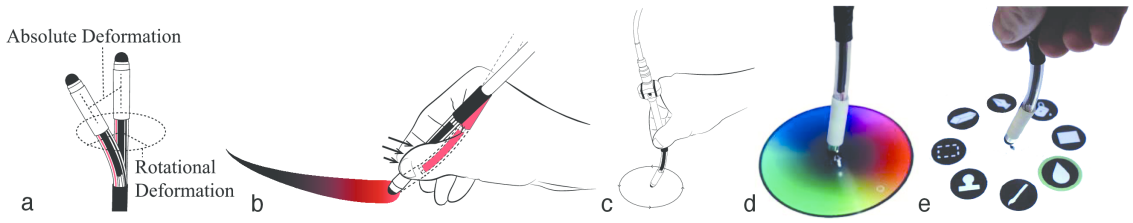

Figure 6. Flexstylus measures absolute and rotational deformation (a). Examples of interaction techniques leveraging bending input: (b) controlling stroke width, (c) performing circle gesture, (d) using color-picker, (e) using a radial menu.

graphical versioning and comparison tool, and rich annotation features that go beyond the text-based tagging commonly employed by existing versioning systems [35]. We plan to further extend this approach with richer annotation capabilities (e.g. media files or examples of use) and query features, and with additional visualizations of the mappings and their changes over time.

7.5.2.3. Probatio: a method and toolkit for generating ideas and prototypes

We have investigated aspects of DMI design by focusing on the complexity of the design space and the importance of prototyping cycles. In particular, we focused on how to provide an initial path for generating DMI ideas and how to reduce the time and effort required to build functional DMI prototypes. We proposed a new methodology and an associated physical prototyping toolkit, *Probatio* (see Fig. 7), which has building blocks inspired by existing instruments [13]. A preliminary user study with musicians and DMI designers revealed a strong potential for its use in the development of DMIs. It also highlighted possible improvements such as more ergonomic shapes for the building blocks and better sound mapping possibilities.

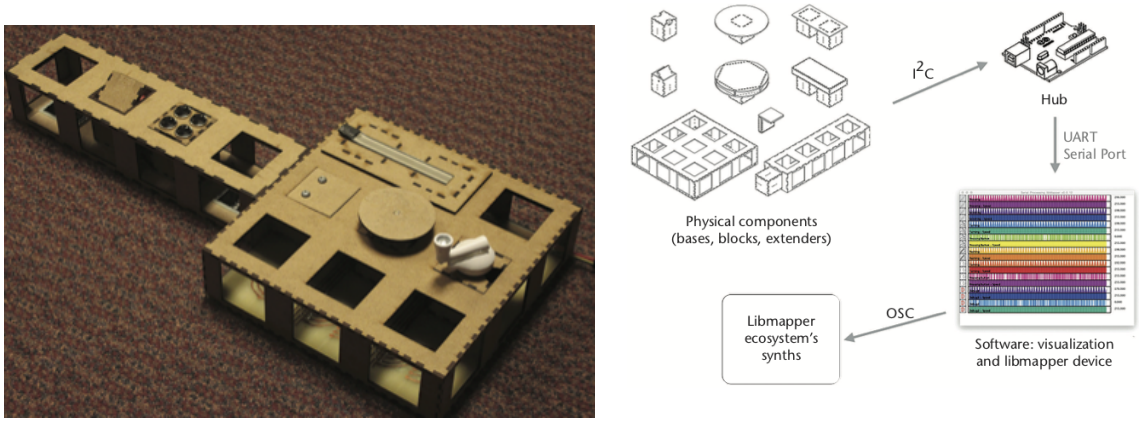

Figure 7. (left) Example of a DMI prototype using Probatio. (right) Overview of Probatio architecture. The blocks are attached to the base, which is connected to an Arduino via I2C. The Arduino is also connected to the computer via serial port. The computer runs an interactive application where each block connected to the base has its value graphed in real time with a unique color, allowing the user to see their activity.