RESEARCH CENTER
Lille - Nord Europe

FIELD

Activity Report 2012

Section New Results

Edition: 2013-04-24

ADAM Project-Team

6. New Results

6.1. Software Product Lines

In terms of Software Product Lines [92], we work in four different directions. First, we define a SPL framework for Cloud Computing called SALOON [62] to face challenges in terms of application configuration, cloud platform configuration [59] and deployment automation [58]. Second, we use Dynamic Software Product Lines (DSLP) for mobile applications [21], in order to support self-adaptation of context-aware applications in ubiquitous environments [56] and Wireless Sensor Networks (WSNs) [36]. In both cases, Constraint Satisfaction Problem (CSP) techniques are used in order to find a suitable configuration for the current environment state and to deal with contradictory dimensions (e.g., accuracy and energy saving) in the decision making process. Third, in the YourCast project [76], we work in a Composite SPL for Broadcasting System by identifying the main issues that we need to deal with when defining such kind of SPL. Finally, we define an operator to compute syntactic and semantic differences between feature models [24].

6.2. Software Evolution

The adaptive software paradigm supports the definition of software systems that are continuously adapted at run-time. An adaptation activates multiple features in the system, according to the current execution context (*e.g.*, CPU consumption, available bandwidth). However, the underlying approaches used to implement adaptation are ordered, *i.e.*, the order in which a set of features are turned on or off matters. Assuming feature definition as etched in stone, the identification of the right sequence is a difficult and time-consuming problem. We propose here a composition operator that intrinsically supports the commutativity of adaptations [50]. Using this operator, one can minimize the number of ordered compositions in a system. It relies on an action-based approach, as this representation can support preexisting composition operators as well as our contribution in an uniform way. This approach is validated on the Service-Oriented Architecture domain, and is implemented using first-order logic.

6.3. Green Middleware

The energy consumption of ICT is widely acknowledged as continuously growing over years and its carbon footprint can now be compared to the avionics domain. While green computing has emerged as a new research area concerned with the optimization of the energy consumption of large-scale systems, such as datacenters, our project-team investigates the analysis of the energy consumption from a software engineering point of view. In particular, we developed e-Surgeon, a middleware framework to estimate the power consumption of legacy software at various levels of granularity. With respect to this objective, the first result we obtained in [52] relates to an evaluation of the impact of programming languages and programming styles on the energy consumption of applications. While the current trend in application servers is to adopt interpreted languages (*e.g.*, JavaScript, Python) on the server side, our preliminary results highlight that these languages impose a large overhead to the energy consumption of the resulting system. In [53], this preliminary result is further investigated by identifying energy hotspots within legacy application servers. To do so, we automatically instrument the code of the application server to analyze how the energy is consumed by the application server under various stress scenarios. Our results show that the energy is mostly consumed by a restricted number of classes and methods of these application servers, thus giving hints to software developers on candidate snippets for optimization.

6.4. Human-Competitive Software Engineering

Frequently asked questions (FAQs) are a popular way to document software development knowledge. As creating such documents is expensive, we have invented an approach for automatically extracting FAQs from sources of software development discussion, such as mailing lists and Internet forums, by combining techniques of text mining and natural language processing. We applied the approach to popular mailing lists and carried out a survey among software developers to show that it is able to extract high-quality FAQs that may be further improved by experts. This research has been published at the International Conference on Software Engineering (ICSE'2012 [40]), the flagship conference in the domain. This work takes place in our line of research on "human competitive software engineering", where we try to replace manual tasks requiring costly human skills (such as documentation writing or bug fixing) by automated or semi-automated approaches.

6.5. Reconfigurable Middleware

In the context of our collaboration with the Thales company, especially via the PhD of Jonathan Labéjof defended on 20 December 2012, we obtained some results in the domain of reconfigurable message-oriented middleware (MOM). MOM are a particular class of middleware systems that promote asynchronous communications and weak coupling between communicating entities. They are of particular interest for the design of Systems of Systems (SoS). In this context, we worked on a method for reconfiguring quality of service properties in MOM. The idea is to be able change the properties of communication channels without stopping these channels. We obtained this by defining a bijection between the characteristics of these channels and a component-based software architecture for which we already have means of reconfiguration with our previous results on the FRASCATI platform (see Section 5.5). By this way, reconfiguring the quality of service of a channel is akin to reconfiguring its associated component-based software architecture. This result has been applied to MOM platforms that conform to the OMG DDS standard.

This result has been the topic of a patent application [106] that was filled in Europe in July 2011 and in the US in July 2012. The results were also presented in the SCDI workshop at the EDOC 2012 conference [44].

ATEAMS Project-Team

5. New Results

5.1. Programming language support for statically type access to external resources

One of the open issues in programming is how to obtain typed access, including its beneficial IDE support, to data sources that have not been modeling with the programming language's data modeling facilities. Rather most data is modeled externally or not modeled at all. Mark Hills, Jurgen Vinju and Paul Klint proposed, designed and validated a programming language design where meta models for external data are imported and/or inferred at compile-time. These models are then used to generate source code to represent these models natively in the idiom of the programming language.

5.2. Statically analyzing PHP code

Tool support in IDEs for PHP code is limited due to the dynamic nature of the language. Mark Hills, Jurgen Vinju and Paul Klint produced a principled yet pragmatic infrastructure for analyzing PHP code nevertheless. The analyses first use crude but effective over-approximations of the PHP semantics to limit the search spaces and improve accuracy, then exploit information from user-manuals, and then use state-of-the-art static analysis techniques in a fixed point abstract interpretation to arrive at accurate results.

5.3. Modular Language Parametric Refactoring Framework

Anastasia Izmaylova with Jurgen Vinju produced a prototype implementation of a framework for specifying refactoring tools based on type constraints that is open to unpredicted language extensions. The problem with the co-evolution of programming language and their supporting refactoring tools is complexity. Often existing refactorings are not retro-fitted with the new language semantics and new opportunities for refactoring tools are not fulfilled. Anastasia designed a solution based on monad transformers that allows the kind of invasive extensibility needed to adapt complex existing implementation of language semantics with additional features that interact in many ways with the existing features.

5.4. Communication Action Emulation

CAE is a novel epistemic model for describing and evaluating the equivalence of communication models by Floor Sietsma and Jan van Eijck.

5.5. Notation-Parametric Grammar Recovery

Vadim Zaytsev generalized the algorithm for recovering context-free grammars from legacy language documentation. This facilitated the recovery of more grammars to be used in the study of grammarware and software language engineering.

5.6. (In)Validating Domain Knowledge Existence in Legacy Source Code

Davy Landman conducted a large experiment in comparing an extensive domain model to the information present in source code of applications that are used in the domain in question. The project management domain was chosen for this. The experiment is still in progress. Big steps were made in setting up the experiment, which includes reporting comprehensively on a large number of design decisions, in a traceable and reproducible manner.

5.7. Ensō

Ensōis a new programming system based on interpretation of domain-specific modeling languages. The system is co-designed and authors by Tijs van der Storm in collaboration with William Cook and Alex Loh. The two foundations of the Ensōsystem are managed data and object grammars. Managed data provides modular strategies for customizing how programming languages represent and provide access to data.

Object grammars form the second foundation: they facilate declarative, compositional, and bidirectional mappings from textual syntax to object graphs. Domain-specific models in Ensōare parsed and rendered using object grammars, and represented, in memory as managed data. Together they combine into a highly flexible and modular platform for model-driven development.

BONSAI Project-Team

6. New Results

6.1. High-throughtput sequence processing

- Within the PhD of T. T. Tran, we proposed a new indexing structure adapted to GPUs. We studied an indexing scheme with perfect hashing functions, and developed a prototype written in openCL for a read mapper. This read mapper has a sensitivity comparable to state-of-the-art read mappers, and provides substantial time gains in some cases.
- Within our collaboration with the Lille hospital on the follow-up of leukemia residual disease, we proposed a new heuristic to study immunological VDJ recombinations and follow their evolution along the time. The method is under testing on several datasets obtained from the Ion Torrent sequencer at IRCL (Institut de Recherche sur le Cancer de Lille).
- Within the PhD of E. Kopylova, we designed an new algorithm to filter out ribosomal RNA sequences from RNA raw data produced in metatranscriptomic sequencing. The method combines text indexing techniques, with the Burst trie, and Universal Levenshtein automaton to allow for seraching with errors. An article has been published the journal *Bioinformatics* [4].

6.2. Noncoding RNAs

- We designed a new algorithm to produce all locally optimal secondary structures of an RNA sequence. Locally optimal secondary structures are thermodynamically stable RNA structures that are maximal for inclusion: they cannot be extended without producing a conflict between base pairs in the secondary structure, or increasing the free energy. This was published in *Journal for Computational Biology* [7].
- We took part to a collaborative work on benchmarking for RNA structure comparison. This work has been published in *Advances in bioinformatics* [2].

6.3. Genomic rearrangements

- Within the context of the PhD of A. Thomas:
  - We designed an algorithm for finding the minimal number of block interchanges required to transform a duplicated linear genome into a tandem duplicated linear genome. We provide a formula for the distance as well as a polynomial time algorithm for the sorting problem. This work was published in the conference *Bioinformatics* [13].
  - We explored a new problem concerning tandem halving, that is reconstructing a non-duplicated ancestor to a partially duplicated genome in a model where duplicated content is caused by several tandem duplications. We provide a distance in $O(n)$ time and a scenario in $O(n^2)$ time. We considered several problems related to multiple tandem reconstruction and proved that the simpliest of reconstructing 2 tandems is NP-hard. This work was published in the conference WABI 2012 [14].
- In the context of ancestral genome reconstruction, we designed an algorithm for the identification of Minimal Conflicting Sets (MCS) rows in a biological binary matrix. We provided a $O(n^2m^2 + nm^7)$ time algorithm, largely improving the up-to-date best algorithm in $O(m^6n^5(m+n)^2log(m+n))$ time. This work was published in the conference CPM 2012 [11].

- In the context of the comparison of sets of alternative gene transcripts, we designed a general framework to compare sets of transcripts that are transcribed from orthologous loci of several species. The model is based on the construction of a common reference sequence, and on annotations that allow the reconstruction of ancestral sequences, the identification of conserved events, and the inference of gains and losses of donor/acceptors sites, exons, introns and transcripts. This work was published in the conference ISBRA 2012 [12].

6.4. Nonribosomal peptides

- With the arrival of Ammar Hasan, a postdoc researcher, we started a new project on the prediction of nonribosomal peptides activity. We defined a novel peptide fingerprint based on monomer composition. This fingerprints is used for peptide similarity searching and for activity prediction. This work was published in *Journal of Computer-Aided Molecular Design* [1].
- We participated in the writing of a review dedicated to kurstakin, a nonribosomal lipopeptide synthetized by several Bacillus genus and published in *Applied microbiology and biotechnology* [3].
- The collaboration with members of EPI Orpailleur (CRI Nancy) succeeded in designing a protocol to discover new nonribosomal peptide synthetases in bacterial genomes and then annotate them in order to predict the peptide they produce. It was published in JOBIM 2012 [16].

DART Project-Team

6. New Results

6.1. Hardware Distributed Control for Dynamic Reconfigurable Systems

The progress in FPGA technology has allowed FPGA-based reconfigurable embedded systems to target increasingly sophisticated applications, which leads to a high design complexity of such systems especially at the adaptation control level. This complexity results into long design phases and delayed time-to-market. In this context, a centralized control model might be not adapted to the growing size and complexity of embedded systems. The use of a single controller for the whole system might result into a high complexity due to the number of parameters to take into account for runtime adaptation, which makes difficult its modification and test. Besides, the design of such a controller is system-dependent since it treats the system as a whole, which represents an obstacle for design reuse. In order to solve these problems, we propose a control design approach aiming to decrease design complexity and enhance design flexibility, reuse and productivity. This approach is based on a semi-distributed control model [34]. In order to achieve the objectives mentioned above, the proposed approach combines autonomy, modularity, formalism and high-level design. The semi-distributed control model divides the control problem between autonomous controllers handling each the self-adaptation of a reconfigurable component of the system, which allows to decrease their design complexity. Each controller handles three main tasks allocated to three different modules: i)monitoring of events that might trigger the adaptation of the controlled component, ii)decision-making about the required adaptations, and iii)adaptation (reconfiguration) realization. To ensure that reconfiguration decisions made by the controllers respect global system constraints such as security and quality of service constraints, these decisions are coordinated before launching the corresponding partial reconfigurations. The allocation of these tasks to separate modules facilitates their modification and reuse and thus the scalability of the control design. For the decision-making modeling, we use the mode-automata formalism. This formalism is suitable to model the control of the different modes of a reconfigurable system such as energy modes or image display modes. Thanks to its clear semantics, the use of such a formalism facilitates the high-level modeling of the controllers and their automatic generation. In order to facilitate code generation and enhance thus design productivity, our control approach makes use of Model-Driven-Engineering (MDE) [33]. Control systems composed of controllers and coordinators are modeled using the UML (Unified Modeling Language) profile MARTE (Modeling and Analysis of Real-Time and Embedded systems). The automation of MDE, allowed to generate the code of these systems. The generated code was then used to validate the semi-distributed control and to determine its resource overhead compared to centralized control systems.

6.2. Regular interconnection network for HP-SoC architecture

Our Synchronous Communication Asynchronous Computation (SCAC) model is a data-parallel execution model dedicated to the High Performance System-on-Chip. The architecture of this model is composed of huge number of complex routers, called node elements (the NEs), communicating and working in perfect synchronizations. Each NE is potentially connected to its neighbors via a regular connection. Furthermore, each NE is connected to a heterogeneous set of computing groups (clusters) allow asynchronous processing. Each group includes a combination of processors programmable, the PEs (software processing units) and specialized hardware accelerators (hardware processing units) to perform critical tasks demanding the more performance. All the system is controlled by a Network Controller Unit, the NCU. The NCU and The PEs are implemented with the Forth processor.

The synchronous communication in SCAC model is presented by two kinds of communications:

- The NCU/NEs communication. In fact, we defined a hNoC model integrated in the SCAC architecture [31]. This model is based on sub-netting the network of processing nodes which separate the control of communication and processing. From this model, our communication system allows a better management of data congestion in the NEs grid through the broadcast with mask of parallel instructions to activated processing nodes.
- The NE/NE communication which is our last contribution. In fact, we defined the X-net interconnection network which is a regular network dedicated to the massively parallel SCAC architecture. This network interconnects directly each PE with its 8 nearest neighbors in a two-dimensional mesh through a specific router in the NE module.

The aim of these last works is to design a regular NoC for SCAC architecture to allow global synchronization of the system communications and increase high performance in terms of area cost and bandwidth. This network based on IP blocks which offer well flexibility and scalability, was implemented in synthesizable VHDL code that was simulated and targeted Xilinx Virtex6 (XC6VLX240T) board. The difficulty of designing X-net is a compromise between an optimal quality of broadcasting, high bandwidth and important flexibility of use, while reducing power consumption and silicon area.

6.3. ReCoMARTE: A Marte Based Profile for Dynamic Reconfigurable Systems Modeling

During the last decade, DPR has been widely studied as a research topic. Despite its intuitive appeal, the technique had eluded widespread adoption, particularly in industrial applications. This is due to the complexities of the provided design flow and the in-depth knowledge of many low level aspects of FPGA technologies used to implement DPR systems. The aim of our current work is to propose a methodology in order to allow us to introduce PDR in MARTE for modeling all types of FPGAs supporting our chosen PDR flow. Afterwards, using the MDE model transformations, the design flow can be used to bridge the gap between high level specifications and low implementation details to finally generate files used by the Xilinx EDK design flow for implementing the top-level SoC description of the system. Indeed, in its current version, UML MARTE profile lacks dynamic reconfiguration concepts and requirements for the reconfiguration at different abstraction levels. We have concentrated our efforts in the creation of the structural description of the system that is used as an input to the DPR design flow to facilitate the design entry phase of the DPR design flow. Therefore, we defined an extended version of MARTE called RecoMARTE (Reconfigurable MARTE) [16] model these concepts mainly at:

- Application level: For reconfigurable applications combining control and data processing, it is very difficult, even impossible to use the MARTE profile for their specification. Non-functional properties such as control concepts are induced by different configurations or running modes of the system and allow the description of more complex behaviours. We recommend a set of extensions to a MARTE profile. We also focus on modelling heterogeneous reconfigurable components, and address the problem of constraints specification for verification issue.
- Control mechanism: We define necessary requirements for the reconfiguration control mechanism in order to manage reconfiguration at every design level. In addition, our solution allows to describe global contracts and constraints for combining automata. Our modeled reconfiguration controller will be then synthesized using Discrete Controller Synthesis formal technique (collaboration work)I n order to always provide a correct configuration to the system, with respect to constraints specified by the designer
- Deployment level: Our design methodology using RecoMARTE enables the deployment, parameterization and integration of hardware IPs into SoC platform at multiple levels of abstraction. We have introduced IP deployment capabilities in MARTE, which aim at facilitating the import of selected low-level features into the high-level models, their modification, and the creation of an IP-XACT design description that is used to parameterize and integrate the underlying IP descriptions.

- Physical level: introduced extensions in MARTE provide some facilities to allow modeling physical architecture of a chosen FPGA. Our solution allows to carry out the physical placement of static and reconfigurable areas on the platform. This task is done through ranges in terms of physical resources, with respect to placement constraints such as consumed resources.

6.4. Using Marte Profile for NoCs modeling

The modeling of repetitive structures such as network on chip topologies in graphics forms poses a particular challenge. This aspect may be encountered in intensive data/control oriented applications such as H.264 video coder. In this work we have described an adequate methodology for modeling NoCs by using the MARTE standard profile. The proposed study has shown that the Repetitive Structure Modeling (RSM) package of MARTE profile is powerful enough for modeling different topologies. By using this methodology, several aspects such as routing algorithm are modeled based finite state machines. This allows to the MARTE profile to be complete enough for modeling a large number of NoCs architectures. Some work is on-going to synthesize such networks in VHDL from such models [55]. While validating the proposed methodology, a co-design approach has been studied by mapping a H264 video coding system onto a Diagonal Mesh NoC by using the Y Chart of Gaspard2 tool. Before allowing the association of the application/architecture, an architectural optimization targeting power minimization of the most critical module of the application and the router of the architecture has been performed. For instance, a flexible VLSI architecture for full-search VBSME (FSVBSME) has been proposed.

6.5. A Hardware Membranes Based Reconfiguration Services Implementation

Partial and dynamic reconfiguration provides a relevant new dimension to design efficient parallel embedded systems. However, due to the encasing complexity of such systems, ensuring the consistency and parallelism management at runtime is still a key challenge. So architecture models and design methodology are required to allow for efficient component reuse and hardware reconfiguration management. We proposed a novel approach inspired from the well-known component based models used in software applications development. Our model is based on membranes wrapping the systems components. The objective is to improve design productivity and ensure consistency by managing context switching and storage using modular distributed hardware controllers. These membranes are distributed and optimized with the aim to design self-adaptive systems by allowing dynamic changes in parallelism degree and contexts migration [26]. These results are obtained in the Famous project by a collaboration with LABSticc Lorient.

6.6. Formal Techniques for General and Domain-Specific Languages

In 2012 we have finished the previous year's activities on domain-specific languages based on formal model-driven engineering with two papers [18], [24]. Our conclusion is that formal MDE-based language definition is interesting because of its generality but adds extra layers of complexity due to the fact that language concepts and semantics are only formalised indirectly, through the formalisation of MDE concepts used in language definition. We have decided thus to move on towards more direct ways of defining and reasoning about languages. We have been experimenting with the K framework [4] for formally defining both the assembly language and a higher-level language for programming on the upcoming dynamically reconfigurable hardware architecture that our team is developing.

We have also worked on proving the correctness of a compiler between high-level and assembly language, based on new symbolic program-equivalence proof techniques we are developping in collaboration with the K team [29]. We have also been working on generic symbolic execution techniques for programming languages having term-rewriting based semantics [28] (PhD of Andrei Arusoaie, supervised in collaboration with Prof. Dorel Lucanu from the K team of Univ. Iasi (Romania).

[4] http://www.k-framework.org

DOLPHIN Project-Team

6. New Results

6.1. On Optimizing a Bi-objective Flowshop Scheduling Problem in Uncertain Environment

Participants: Arnaud Liefooghe, Laetitia Jourdan, El-Ghazali Talbi

Existing models from scheduling often over-simplify the problems appearing in real-world industrial situations. The original application is often reduced to a single-objective one, where the presence of uncertainty is neglected. In [23], we focus on multi-objective optimization in uncertain environments. A bi-objective flowshop scheduling problem with uncertain processing times is considered. An indicator-based evolutionary algorithm is proposed to handle these two difficulties (multiple objectives and uncertain environment) at the same time. Four different strategies, based on uncertainty-handling quality indicators, are proposed in the paper. Computational experiments are performed on a large set of instances by considering different scenarios with respect to uncertainty. We show that an uncertainty-handling strategy is a key issue to obtain good-quality solutions, and that the algorithm performance is strongly related to the level of uncertainty over the environmental parameters.

6.2. New Price settings models in the energy field

Participants: L. Brotcorne, S. Afsar

The electricity supply industry is facing in many countries a restructuring process towards deregulation and competition. In that context classical marginal cost based approaches based on estimation of cost production function and demand functions are not well-suited anymore. Indeed, the energy prices have to be defined not only to retrieve the production costs but also in order to take into account the consumer behavior. Consumers make their choice of service, or of energy provider in order to minimize their disutility values. Failing to recognize that may lead to tremendous lack on revenues. In order to capture this hierarchical decision process where a leader (the energy provider) takes explicitly into account the reaction of a follower (the consumers) in his decision process. The energy pricing problems addressed in this are modeled as bilevel programs.

6.3. Bi-level formulation for a Long-Distance Freight Transportation Problem

Participants: M. Diaby, L. Brotcorne, E.-G. Talbi.

A company wants to convey different types of products from origin i to points of destination j. It can deliver the goods itself or hire a transport company, and subcontract part of the application. The transport company must offer attractive prices while aiming to maximize its profit. The aim of this problem is to determine rates that allow the carrier to maximize its revenues and remain affordable for the customer. The problem is modeled as a bilevel program at the first level, the carrier (leader) wants to maximize its revenues; at the second level, the client. An exact and an evolutionary solution approaches are developped.

6.4. On Local Search for Bi-objective Knapsack Problems

Participant: A. Liefooghe.

In [26], a local search approach is proposed for three variants of the bi-objective binary knapsack problem, with the aim of maximizing the total profit and minimizing the total weight. First, an experimental study on a given structural property of connectedness of the efficient set is conducted. Based on this property, a local search algorithm is proposed and its performance is compared against exact algorithms in terms of running time and quality metrics. The experimental results indicate that this simple local search algorithm is able to find a representative set of optimal solutions in most of the cases, and in much less time than exact algorithms.

6.5. Convergent methods based on aggregation in mathematical models

Participant : François Clautiaux

We designed several algorithms to aggregate variables in integer linear programs. Our methods first solve aggregated models, and converge to the optimal solution of the initial problem by iteratively refining the model.

The first method applies on a large network flow models that use a pseudo-polynomial number of variables. It is based on an initial aggregation of the vertices of the model and its iterative refinement using different optimization techniques. This led to dramatical improvements for a special case of vehicle routing problem. We proposed several theoretical results regarding convergence, suitable discretizations, wort-case analysis and approximation algorithms [44].

The second method applies on column generation approaches for the cutting-stock problem. Our algorithm links groups of dual variables by linear constraints, leading to a problem of smaller dimension, whose solutions are dual-feasible for the initial problem. The corresponding "inner approximation" is iteratively refined by splitting the groups into smaller groups until an optimal dual solution is found. This method allows to produce a valid lower bound at each iteration, which is not the case for classical column-generation schemes [58].

6.6. Investigating the Optimization Goal of Indicator-Based Multiobjective Search

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Heike Trautmann and Tobias Wagner (TU Dortmund University, Germany)

Using a quality indicator in the environmental selection step of evolutionary multiobjective optimization (EMO) algorithms to indicate which solutions shall be kept in the algorithms' population and which should be deleted, introduces a certain search bias. Instead of an "arbitrary" subset of the Pareto front, such (quality) indicator based search algorithms aim at approximating the set of μ solutions that optimizes a given indicator, for which the term *optimal μ-distribution* has been introduced [63]. Also for performance assessment with respect to a given indicator, knowledge about the optimal μ-distributions is helpful as interpreting the *achieved* indicator values with respect to the best *achievable* value becomes possible. For the hypervolume indicator, several results on these optimal μ-distributions are known [63], [62], [75], [69], [70], [61] [64], but the understanding of the optimization goal for other indicators is less developed. Recently, we started to investigate the optimal μ-distributions, both theoretically and numerically, for the so-called $R2$ indicator [79]—another often recommended quality indicator [90]. Instead of the binary version of [79] that takes two solution sets and assigns them a certain quality, we thereby investigated an equivalent unary indicator where one (reference) set is always fixed.

First experiments for problems with two objectives and connected Pareto fronts have been presented in [37] which won the best paper award within the EMO track at GECCO'2013 [5]. Further investigations on problems with disconnected Pareto fronts have been submitted to the Evolutionary Computation journal [72]. We also studied in more detail how the parameters of the $R2$ indicator such as the ideal point or the distribution of weight vectors can be used to change the optimization goal [86] and correspondingly proposed the algorithm $R2$-EMOA which is able to steer the search towards preferred regions of the Pareto front by optimizing the $R2$ indicator directly in its environmental selection [85], [72].

6.7. Runtime Analyses of Interactive Evolutionary Algorithms

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Manuel López-Ibáñez (Université Libre de Bruxelles, Belgium), Boris Naujoks (Cologne University of Applied Sciences, Germany), and Gunter Rudolph (TU Dortmund University, Germany)

[5] See http://www.sigevo.org/gecco-2012/papers.html.

If a decision maker (DM) expresses preferences, e.g., towards certain points or regions of the search space, during the algorithm run, we call such an algorithm *interactive*. Interactive algorithms are frequently used in the field of multi-criteria decision making, but theoretical results on interactive evolutionary multiobjective algorithms (EMOAs) have not been derived until recently. In [36], we started to analyze interactive versions of an evolutionary algorithm with plus-selection and a population size of one, the so-called iRLS and $i(1+1)$EA. On two pseudo-boolean problems, recently used for theoretically analyzing EMOAs, we could prove upper bounds on the expected runtime of the two mentioned algorithms and on the number of times, the DM is asked about his/her preferences until the most-preferred search point is found. The analyzes showed that the internal value function of the DM has a strong, non-desired influence on the algorithms' runtimes and that the number of questions to the DM are too high for a practical relevant algorithm. It is an open question which algorithm designs are necessary to circumvent these two drawbacks.

6.8. Benchmarking of CMA-ES Variants for Numerical Blackbox Optimization

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Anne Auger and Nikolaus Hansen (Inria Saclay - Ile-de-France)

The covariance matrix adaptation evolution strategy (CMA-ES) is one of the state-of-the-art optimization algorithms for numerical single-objective blackbox optimization [81], [80] [67]. Previously, we proposed to use so-called mirrored mutations to generate new candidate solutions in evolution strategies which turned out to increase the convergence rate for certain variants [71], [65], [66]. Another recent approach to speed up the CMA-ES is to perform an active (i.e. negative) covariance matrix update [60]. In [32], [35], [34], [33], we tested empirically how the combination of mirrored mutations and active CMA-ES perform on the COCO framework [77], [78]. It turned out that both concepts complement each other well without a significant decrease in performance on any of the 24 test functions. Moreover, the main improvement over the standard CMA-ES could be shown to come from the active covariance matrix adaptation while the addition of mirrored mutations only slightly improves the algorithm.

6.9. Self-adaptive method for a three-objective vector-packing problem

Participants: Nadia Dahmani, François Clautiaux, El-Ghazali Talbi

We introduced a new multi-objective packing problem (MOBPP), in which we optimize the number of bins, the maximum weight of a bin, and the loading balancing. We studied the impact and the combination of two complementary decoding strategies for this problem. A feature of our work is to insert the parameters of the decoders in the representation of the solution. It leads to self-adaptive meta-heuristics, where the algorithm iteratively adapts the parameters during the search. We embedded our approaches in a local search and an evolutionary algorithm for the MOBPP. A comprehensive set of experiments were performed on various benchmarks inspired from the literature. Results confirm that our methods lead to more effective muti-objective metaheuristics for this problem.

6.10. Multiple Neighborhood Exploration Through Adaptive Search

Participants: Bilel Derbel, Houda Derbel, El-Ghazali Talbi, Hiba Yahyaoui

Variable neighborhood descent (VND) and its several variants are based on the systemic change of neighborhoods within the search. It is well known that the performance of a VND-like algorithm highly depends on the order/way the neighborhoods are alternated. In this work, we focus on designing new meta-strategies for deciding what neighborhood structure to apply through the search. Two new approaches are proposed to tackle this issue. In the first approach [41], we model the search by considering the neighborhood tree induced by alternating the use of different structures within a local search descent. We investigate the issue of designing a search strategy operating at the neighborhood tree level by exploring different paths of the tree in a heuristic way. We show that allowing the search to 'backtrack' to a previously visited solution and resuming the iterative

variable neighborhood descent by 'pruning' the already explored neighborhood branches leads to the design of effective and efficient search heuristics. In the second approach, we investigate deterministic and randomized adaptive strategies for selecting the next neighborhood to apply at runtime. The adaptive strategies are based on computing a reward for each neighborhood with respect to the observed average ratio of solution quality and time cost.

6.11. CoBRA: A cooperative coevolutionary algorithm for bi-level optimization

Participants: François Legillon, Arnaud Liefooghe, El-Ghazali Talbi

In [43] we present CoBRA, a new evolutionary algorithm, based on a coevolutionary scheme, to solve bi-level optimization problems. It handles population-based algorithms on each level, each one cooperating with the other to provide solutions for the overall problem. Moreover, in order to evaluate the relevance of CoBRA against more classical approaches,a new performance assessment methodology, based on rationality, is introduced. An experimental analysis is conducted on a bi-level distribution planning problem, where multiple manufacturing plants deliver items to depots, and where a distribution company controls several depots and distributes items from depots to retailers. The experimental results reveal significant enhancements, particularly over the lower level, with respect to a more classical approach based on a hierarchical scheme.

6.12. Neutrality in the Graph Coloring Problem

Participants: Marie-Eléonore Marmion, Aymeric Blot, Laetitia Jourdan, and Clarisse Dhaenens

The graph coloring problem is often investigated in the literature. Many insights about the existence of many neighboring solutions with the same fitness value are raised but as far as we know, no deep analysis of this neutrality has ever been conducted in the literature. We have quantified the neutrality of some hard instances of the graph coloring problem. This neutrality property has to be detected as it impacts the search process. Indeed, local optima may belong to plateaus that represents a barrier for local search methods. We also aim at pointing out the interest of exploiting neutrality during the search. Therefore, a generic local search dedicated to neutral problems (NILS) and previously tested on flowshop problems, is performed and tested on several hard instances. Results show that taking into account neutrality allows to obtain better results than when not considering it.

6.13. Local Search in the Context of Classification Rule Mining

Participants: Julie Jacques, Laetitia Jourdan, Clarisse Dhaenens

Many multi-objective algorithms have been proposed to solve the classification rule mining problem; the vast majority of them are based on genetic algorithms. We propose an algorithm, MOCA - Multi-Objective Classification Algorithm -, to solve this problem. The originality of MOCA is to be a dominance-based multi-objective local search (DMLS) using a Pittsburgh representation of rules. We evaluated several DMLS implementations and neighborhood operators on literature datasets and one real dataset. Then we compared the best obtained algorithm against several efficient approaches of the literature. The experiments show that the proposed approach is very competitive in comparison to other algorithms tested. Moreover, our approach is able to deal with very large real datasets and manages to have a good accuracy.

6.14. MOCA-I: discovering rules and guiding decision maker in the context of partial classification in large and imbalanced datasets

Participants: Julie Jacques, Laetitia Jourdan, Clarisse Dhaenens

In this work we focus on the modeling and the implementation as a multi-objective optimization problem of a Pittsburgh classification rule mining algorithm adapted to large and imbalanced datasets, as encountered in hospital data. Indeed hospital data comes with problems such as class imbalance, volumetry or inconsistency, and optimization approaches have to take into account such specificities. We present MOCA-I, an adaptation of MOCA dedicated to this kind of problems. We propose its implementation as a dominance-based local search in opposition to existing multi-objective approaches based on genetic algorithms. We associate to this algorithm an original post-processing method based on the ROC curve to help the decision maker to choose the most interesting set of rules. Our approach is currently compared to state-of-the-art classification rule mining algorithms (both classic approaches such as C4.5 and optimization approaches), giving as good or better preliminary results, using less parameters. Moreover, our approach has been compared to C4.5 and C4.5-CS on a real dataset (hospital data) with a larger set of attributes, giving the best results. The complete evaluation is still going on.

6.15. A method to combine combinatorial optimization and statistics to mine high-throughput genotyping data

Participants : Julie Hamon, Clarisse Dhaenens, Julien Jacques (MODAL)

In the context of genomic analysis (collaboration with Genes Diffusion), dealing with high-throughput genotyping data, the objective of our study is to select a subset of SNPs (single nucleotide polymorphisms) explaining a trait of interest. We propose a method combining combinatorial optimization and statistics to extract a subset of interesting SNPs. The combinatorial part aims at exploring in an efficient way the large search space induced by the large number of possible subsets and statistics are used to evaluate the selection. We propose a first method based on an ILS (iterated local search) and using a regression. Three criteria used to evaluate the quality of the regression are compared. One of them (the k-fold validation) shows better performance. We also compare this approach to classical statistical approaches on simulated datasets. Results are promising as the proposed approach outperforms most of these statistical approaches [51].

6.16. Design and implementation of performance or energy-aware parallel optimization algorithms

Problems in practice are nowadays becoming more and more complex and time-intensive and their resolution requires to harness more and more computational resources. In parallel, the recent advances in hardware architecture enable to provide such required tremendous computational power through massively multi-core and GPU infrastructures. Such huge amount of cores is often provided through heterogeneous single or multi-clusters. The exploitation of such infrastructures clearly poses two fundamental and conflictual issues which are two major challenging perspectives of the Dolphin project that have been investigated during the 2012 year: (1) *Performance-aware issue*: how to design, implement and validate efficient and effective algorithms for such target machines to solve large size combinatorial optimization problems? ; (2) *Energy-aware issue*: using a large amount of computational resources for the deployment of large scale algorithms is energy-consuming. Therefore, the second issue is how to deal with the performance issue with a minimized cost in terms of energy consumption?
To deal with these issues, we have proposed new approaches summarized in the following sections.

6.16.1. Design and implementation of performance-aware optimization algorithms

In order to allow one to solve large size combinatorial optimization problems, we have revisited the design and implementation of meta-heuristics and exat (B&B) algorithms for two major hardware platforms: heterogeneous multi and many-core clusters and computational grids including multiple clusters.

- **Multi-core GPU-based hybrid meta-heuristics** - *Participants: T-V. Luong, N. Melab and E-G. Talbi.*

In [28], we have revisited the design and implementation of respectively single-solution and population-based meta-heuristics for single-core CPU coupled with a GPU device. We have investigated and proposed a new guidline for combining multi-core and GPU computing for hybrid meta-heuristics. Efficient approaches have been proposed for CPU-GPU data transfer optimization and task repartition between the GPU device and the CPU cores. Extensive experiments have been performed on an 8-core CPU coupled with a GPU card using Ant colonnies combined with a local serach applied to the Quadratic Assignment Problem (QAP). The reported results show that the use of multi-core computing, in addition to GPU computing, provides a performance improvement of up to 75%.

- **GPU-accelerated Branch-and-Bound algorithms** - *Participants: I. Chakroun and N. Melab.*
  Branch-and-Bound (B&B) algorithms are based on an implicit enumeration of a dynamically built tree-based search space. Pruning tree nodes (sub-problems) is traditionally used as a powerful mechanism to reduce the size of the explored search space. Such mechanism requires to perform the bounding operation which consists in applying a lower bound function to the generated sub-problems. Preliminary experiments performed on the Flow-Shop scheduling problem (FSP) have shown that the bounding operation consumes over 98% of the execution time of the B&B algorithm. Therefore, we have investigated the use of GPU computing as a major complementary way to speed up the search. We have revisited the design and implementation of the parallel bounding model for FSP on GPU accelerators dealing with two major issues: (1) thread divergence caused by the highly irregular nature of the explored tree and the SIMD execution model of GPU ; (2) data access optimization required for mapping efficiently different data structures on the hierarchy of memories provided in the GPU device. In [14], we have proposed a GPU-based parallel bounding model together with a data refactoring approach to deal with thread divergence. In [45] (an extended version submitted to the CCPE journal is being revised), we have proposed an efficient data optimization strategy based on a deep analysis of the complexity of the different data structures of the FSP lower bound algorithm in terms of memory size and access latency. The different proposed approaches for the two issues have been extensively experimented using and Nvidia Tesla C2050 GPU card. Compared to a CPU-based execution, accelerations up to more than $\times 100$ are achieved for large problem instances.
- **Peer-to-peer Branch-and-Bound algorithms** - *Participants: T-T. Vu, B. Derbel and N. Melab.*
  To deal with dynamic load balancing in large scale distributed systems, we have proposed in [50] to organize computing resources following a logical peer-to-peer overlay and to distribute the load according to the so-defined overlay. We have used a tree as a logical structure connecting distributed nodes and we balance the load according to the size of induced subtrees. We have conducted extensive experiments involving up to 1000 computing cores and provided a throughout analysis of different properties of our generic approach for two different applications, namely, the standard Unbalanced Tree Search and the more challenging parallel Branch-and-Bound algorithm. Substantial improvements are reported in comparison with the classical random work stealing and two finely tuned application specific strategies taken from the literature.

6.16.2. Design and implementation of energy-aware optimization algorithms

- Participants: Y. Kessaci, N. Melab and E-G. Talbi.

Cloud computing is an emerging computer science paradigm of distributed computing in which applications, data and infrastructures are proposed as a service that can be consumed in a ubiquitous, flexible and transparent way. Cloud computing brings with it such benefits via cloud managers which hide to the user some complex and challenging issues such as scheduling. However, the solutions to these issues provided in cloud managers are sometimes limited. For instance, the scheduling approach proposed in many cloud managers like OpenNebula is limited regarding the criteria taken into account. Energy consumption, which is highly critical for many applications such as High Performance Computing (HPC), is rarely considered. In [42] (selected for a special issue of FGCS journal) , we have addressed energy-aware scheduling of energy and time-consuming applications for cloud infrastructures. We have proposed a multi-start parallel

local search heuristic for cloud managers (EMLS-ONC) with the focus put on OpenNebula. EMLS-ONC has been experimented using different (VMs) arrival scenarii and different hardware infrastructures. The results show that EMLS-ONC outperforms the scheduler provided in OpenNebula by a significant margin in terms of energy consumption and number of scheduled VMs.

FUN Team

5. New Results

5.1. Routing in FUN

Participants: Nicolas Gouvy, Xu Li, Nathalie Mitton.

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 geographic routing paradigms. It first has considered a static network in which the routing either enhances the energy cost [22], [10], balances the load over nodes [21], [8] or respects traffic priorities [18].

A more complex routing paradigm has been proposed in [25] for k-anycasting. In k-anycasting, a sensor wants to report event information to any k sinks in the network. This is important to gain in reliability and efficiency in wireless sensor and actor networks. In this paper, we describe KanGuRou, the first position-based energy efficient k-anycast routing which guarantees the packet delivery to k sinks as long as the connected component that contains s also contains sufficient number of sinks. A node s running KanGuRou first computes a tree including k sinks among the M available ones, with weight as low as possible. If this tree has $m \geq 1$ edges originated at node s, s duplicates the message m times and runs m times KanGuRou over a subset of defined sinks. Simulation results show that KanGuRou allows up to 62% of energy saving compared to plain anycasting.

We then assumed that the sink that collects data is actually mobile and travels the network. Sensor nodes need thus to update the position of the sink in a smart fashion in order to limit the overhead generated by this update. In [9], we propose a novel localized Integrated Location Service and Routing (ILSR) scheme, based on the geographic routing protocol GFG, for data communications from sensors to a mobile sink in wireless sensor networks. The objective is to enable each sensor to maintain a slow-varying routing next hop to the sink rather than the precise knowledge of quick-varying sink position. In ILSR, sink updates location to neighboring sensors after or before a link breaks and whenever a link creation is observed. Location update relies on flooding, restricted within necessary area, where sensors experience (next hop) change in GFG routing to the sink. Dedicated location update message is additionally routed to selected nodes for prevention of routing failure. Considering both unpredictable and predictable (controllable) sink mobility, we present two versions. We prove that both of them guarantee delivery in a connected network modeled as unit disk graph. ILSR is the first localized protocol that has this property. We further propose to reduce message cost, without jeopardizing this property, by dynamically controlling the level of location update. A few add-on techniques are as well suggested to enhance the algorithm performance. We compare ILSR with an existing competing algorithm through simulation. It is observed that ILSR generates routes close to shortest paths at dramatically lower (90% lower) message cost.

When the network is composed of mobile sensors that have the faculty to control their mobility, this property can be exploited to enhance routing performance. In [3], we are interested in energy-aware routing algorithms that explicitly take advantage of node mobility to improve energy consumption of computed paths. Mobility is a two-sword edge however. Moving a node may render the network disconnected and results in early termination of information delivery. To mitigate these problems, we propose a family of routing algorithm called CoMNet (Connectivity preservation Mobile routing protocols for actuator and sensor NETworks), that uses local information and modifies the network topology to support resource efficient transmissions. Our extensive simulations show that CoMNet has high energetic performance improvement compared to existing routing algorithms. More importantly, we show that CoMNet guarantees network connectivity and efficient resource consumption.

5.2. Self-organization

Participants: Tony Ducrocq, Xu Li, Nathalie Mitton.

Self-organization encompasses several mechanisms [35]. This year, the FUN research group contributes to some of them such as neighbor discovery, localization, clustering and topology control in FUN.

5.2.1. Neighbor discovery

To perform routing or any specific task, a node needs to discover its neighbors. Hello protocol is the basic technique for neighborhood discovery in wireless ad hoc networks. It requires nodes to claim their existence/aliveness by periodic 'hello' messages. Central to a hello protocol is the determination of hello message transmission rate. No fixed optimal rate exists in the presence of node mobility. The rate should in fact adapt to it, high for high mobility and low for low mobility. In [31], we combine parameters of the neighborhood discovery (sending frequency of hello messages and changes in the neighborhood tables) and transmission range of the nodes. We present two algorithms that adapt transmission range of the sensors in a mobile WSN by still adapting frequency of hello messages in order to save energy and get accurate neighborhood tables. The first solution is based on the knowledge of turnover - change in the number of neighbors in consecutive iterations of the neighborhood discovery - used in conjunction with an adaptation of the message frequency and the transmission range, minimizing overall transmission cost of hello messages. The second solution is based on the computation of optimal range knowing the nodes' speed. Both algorithms are based on theoretical analysis. Results show that they are energy efficient and outperform solutions of the literature by maintaining high accuracy.

5.2.2. Topology control

Topology control is a tool for self-organizing wireless networks locally. It allows a node to consider only a subset of links/neighbors in order to later reduce computing and memory complexity. Topology control in wireless sensor networks is an important issue for scalability and energy efficiency. It is often based on graph reduction performed through the use of Gabriel Graph or Relative Neighborhood Graph. This graph reduction is usually based on geometric values.

In [7], we propose a radically new family of geometric graphs, i.e., Hypocomb, Reduced Hypocomb and Local Hypocomb for topology control. The first two are extracted from a complete graph; the last is extracted from a Unit Disk Graph (UDG). We analytically study their properties including connectivity, planarity and degree bound. All these graphs are connected (provided the original graph is connected) planar. Hypocomb has unbounded degree while Reduced Hypocomb and Local Hypocomb have maximum degree 6 and 8, respectively. To our knowledge, Local Hypocomb is the first strictly-localized, degree-bounded planar graph computed using merely 1-hop neighbor position information. We present a construction algorithm for these graphs and analyze its time complexity. Hypocomb family graphs are promising for wireless ad hoc networking. We report our numerical results on their average degree and their impact on FACE [39] routing. We discuss their potential applications and some open problems.

5.2.3. Localization

In mobile-beacon assisted sensor localization, beacon mobility scheduling aims to determine the best beacon trajectory so that each sensor receives sufficient beacon signals with minimum delay. We propose a novel DeteRministic bEAcon Mobility Scheduling (DREAMS) algorithm [6], without requiring any prior knowledge of the sensory field. In this algorithm, beacon trajectory is defined as the track of depth-first traversal (DFT) of the network graph, thus deterministic. The mobile beacon performs DFT under the instruction of nearby sensors on the fly. It moves from sensor to sensor in an intelligent heuristic manner according to RSS (Received Signal Strength)-based distance measurements. We prove that DREAMS guarantees full localization (every sensor is localized) when the measurements are noise-free. Then we suggest to apply node elimination and topology control (Local Minimum Spanning Tree) to shorten beacon tour and reduce delay. Through simulation we show that DREAMS guarantees full localization even with noisy distance measurements. We evaluate its performance on localization delay and communication overhead in comparison with a previously proposed static path based scheduling method.

5.2.4. Clustering

Clustering in wireless sensor networks is an efficient way to structure and organize the network. It aims to identify a subset of nodes within the network and bind it a leader (i.e. cluster-head). This latter becomes in charge of specific additional tasks like gathering data from all nodes in its cluster and sending them by using a longer range communication to a sink. As a consequence, a cluster-head exhausts its battery more quickly than regular nodes. In [14], we present BLAC, a novel Battery-Level Aware Clustering family of schemes. BLAC considers the battery-level combined with another metric to elect the cluster-head. It comes in four variants. The cluster-head role is taken alternately by each node to balance energy consumption. Due to the local nature of the algorithms, keeping the network stable is easier. BLAC aims to maximize the time with all nodes alive to satisfy application requirements. Simulation results show that BLAC improves the full network lifetime 3-time more than traditional clustering schemes by balancing energy consumption over nodes and still delivering high data percentage.

5.3. Self-deployment

Participants: Milan Erdelj, Xu Li, Karen Miranda, Enrico Natalizio, Tahiry Razafindralambo, Dimitris Zorbas.

Robot self-deployment may have different purposes. The FUN research group has addressed four of them that are *(i)* area coverage, *(ii)* barrier coverage, *(iii)* point of interest coverage and *(iv)* deployment for substitution networks.

5.3.1. Area coverage

In [1], with the focus on the self-organizing capabilities of nodes in WSRN, we propose a movement-assisted technique for nodes self-deployment. Specifically, we propose to use a neural network as a controller for nodes mobility and a genetic algorithm for the training of the neural network through reinforcement learning. This kind of scheme is extremely adaptive, since it can be easily modified in order to consider different objectives and QoS parameters. In fact, it is sufficient to consider a different kind of input for the neural network to aim for a different objective. All things considered, we propose a new method for programming a WSRN and we show practically how the technique works, when the coverage of the network is the QoS parameter to optimize. Simulation results show the flexibility and effectiveness of this approach even when the application scenario changes (e.g., by introducing physical obstacles).

In [4], we tackle the issue in a different way. We leverage prediction by exploiting temporal-spatial correlations among sensory data. The basic idea lies in that a sensor node can be turned off safely when its sensory information can be inferred through some prediction methods, like Bayesian inference. We adopt the concept of entropy in information theory to evaluate the information uncertainty about the region of interest (RoI). We formulate the problem as a minimum weight sub-modular set cover problem, which is known to be NP hard. To address this problem, an efficient centralized truncated greedy algorithm (TGA) is proposed. We prove the performance guarantee of TGA in terms of the ratio of aggregate weight obtained by TGA to that by the optimal algorithm. Considering the decentralization nature of WSNs, we further present a distributed version of TGA, denoted as DTGA, which can obtain the same solution as TGA. The implementation issues such as network connectivity and communication cost are extensively discussed. We perform real data experiments as well as simulations to demonstrate the advantage of DTGA over the only existing competing algorithm and the impacts of different parameters associated with data correlations on the network lifetime.

In [34], [13], we leverage some assumptions. One of the main operations in wireless sensor networks is the surveillance of a set of events (targets) that occur in the field. In practice, a node monitors an event accurately when it is located closer to it, while the opposite happens when the node is moving away from the target. This detection accuracy can be represented by a probabilistic distribution. Since the network nodes are usually randomly deployed, some of the events are monitored by a few nodes and others by many nodes. In applications where there is a need of a full coverage and of a minimum allowed detection accuracy, a single node may not be able to sufficiently cover an event by itself. In this case, two or more nodes are needed to collaborate and to cover a single target. Moreover, all the nodes must be connected with a base station that

collects the monitoring data. In this paper we describe the problem of the minimum sampling quality, where an event must be sufficiently detected by the maximum possible amount of time. Since the probability of detecting a single target using randomly deployed static nodes is quite low, we present a localized algorithm based on mobile nodes. Our algorithm sacrifices a part of the energy of the nodes by moving them to a new location in order to satisfy the desired detection accuracy. It divides the monitoring process in rounds to extend the network lifetime, while it ensures connectivity with the base station. Furthermore, since the network lifetime is strongly related to the number of rounds, we propose two redeployment schemes that enhance the performance of our approach by balancing the number of sensors between densely covered areas and areas that are poorly covered. Finally, our evaluation results show an over 10 times improvement on the network lifetime compared to the case where the sensors are static. Our approaches, also, outperform a virtual forces algorithm when connectivity with the base station is required. The redeployment schemes present a good balance between network lifetime and convergence time.

5.3.2. Barrier coverage

Barrier coverage problem in emerging mobile sensor networks has been an interesting research issue. Existing solutions to this problem aim to decide one-time movement for individual sensors to construct as many barriers as possible, which may not work well when there are no sufficient sensors to form a single barrier. In [19], we try to achieve barrier coverage in sensor scarcity case by dynamic sensor patrolling. In specific, we design a periodic monitoring scheduling (PMS) algorithm in which each point along the barrier line is monitored periodically by mobile sensors. Based on the insight from PMS, we then propose a coordinated sensor patrolling (CSP) algorithm to further improve the barrier coverage, where each sensor's current movement strategy is decided based on the past intruder arrival information. By jointly exploiting sensor mobility and intruder arrival information, CSP is able to significantly enhance barrier coverage.We prove that the total distance that the sensors move during each time slot in CSP is the minimum. Considering the decentralized nature of mobile sensor networks, we further introduce two distributed versions of CSP: S-DCSP and G-DCSP. Through extensive simulations, we demonstrate that CSP has a desired barrier coverage performance and S-DCSP and G-DCSP have similar performance as that of CSP.

5.3.3. Point of Interest coverage

The coverage of Points of Interest (PoI) is a classical requirement in mobile wireless sensor applications. Optimizing the sensors self-deployment over a PoI while maintaining the connectivity between the sensors and the base station is thus a fundamental issue. This algorithm addresses the problem of autonomous deployment of mobile sensors that need to cover a predefined PoI with a connectivity constraint. In our algorithm [2], each sensor moves toward a PoI but has also to maintain the connectivity with a subset of its neighboring sensors that are part of the Relative Neighborhood Graph (RNG). The Relative Neighborhood Graph reduction is chosen so that global connectivity can be provided locally. Our deployment scheme minimizes the number of sensors used for connectivity thus increasing the number of monitoring sensors. Analytical results, simulation results and practical implementation are provided to show the efficiency of our algorithm.

We then extended this coverage to multiple points of interest in [15], [16]. Indeed, the problems of multiple PoI coverage, environment exploration and data report are still solved separately and there are no works that combine the aforementioned problems into a single deployment scheme. In this work, we have extended [2] to multiple PoI coverage and combined it to and environment exploration in order to capture the dynamics of the monitored area. We examine the performance of our scheme through extensive simulation campaigns.

5.3.4. Substitution networks

A substitution network is a temporary network that will be deployed to support a base network in trouble and help it to provide best service. [11], [24] present how the mobility of routers impacts the performance of a wireless substitution network. To that end, we simulate a scenario where a wireless router moves between three static nodes, a source and two destinations of UDP traffic. Specifically, our goal is to deploy or redeploy the mobile relays so that application-level requirements, such as data delivery or latency, are met. Our proposal for a mobile relay achieves these goals by using an adaptive approach to self-adjust their position based

on local information. We obtain results on the performance of end-to-end delay, jitter, loss percentage, and throughput under such mobility pattern for the mobile relay. We show how the proposed solution is able to adapt to topology changes and to the evolution of the network characteristics through the usage of limited neighborhood knowledge.

5.4. MAC layer

Participant: Tahiry Razafindralambo.

Multihop wireless networks are used to provide Internet connectivity to the users and the level of performance and quality expected by these users are increasing. In order to meet these performance and quality requirements, wireless communications should be enhanced. Previous works from the literature show that the performance and quality provided by an IEEE 802.11-based multihop wireless network are far from optimal and that there exist different ways to increase the efficiency and the quality of service of such a network. Some studies show that using the medium state as a parameter to tune the behavior of an IEEE 802.11-based multihop network is an appropriate way to proceed. A station in a IEEE 802.11-based multihop wireless network senses the medium either busy or idle. The durations of idle periods and busy periods and their distributions have a clear impact on the network and nodes performance. The understanding of the relationship between these indicators, namely idle and busy periods, the network topology and the traffic, would give new insights to enhance the performance and quality of multihop wireless networks. Due to its multihop and distributed nature, the characterization of idle period durations is difficult in such a network. In [27], [26], we explore the characterization of idle period distribution by proposing a new analytical model and provides an application of this characterization with the design of an adaptive backoff algorithm based on idle periods.

5.5. Servicing

Participants: Xu Li, Kalypso Magklara, Nathalie Mitton, Tahiry Razafindralambo, Dimitris Zorbas.

Servicing wireless sensor networks include many primitives. It can range from cloud connection [12] to mobile IPv6 management [29] going through energy prediction [20] and launching mobile robots on request of a specific demand [5] or to reload sensors [23], [17].

5.5.1. Node reloading

A critical problem of wireless sensor networks is the network lifetime, due to the device's limited battery lifetime. The nodes are randomly deployed in the field and the system has no previous knowledge of their position. To tackle this problem, in [23], we use a mobile robot, that discovers the nodes around it and replaces the active nodes, whose energy is drained, by fully charged inactive nodes. We propose two localized algorithms, that can run on the robot and that decide, which nodes to replace. We simulate our algorithms and our findings show that all nodes that fail are replaced in a short period of time.

In [17] we focus on an emerging kind of cooperative networking system in which a small team of robotic agents lies at a base station. Their mission is to service an already-deployed WSN by periodically replacing all damaged sensors in the field with passive, spare ones so as to preserve the existing network coverage. This novel application scenario is here baptized as "multiple-carrier coverage repair" (MC2R) and modeled as a new generalization of the vehicle routing problem. A hybrid metaheuristic algorithm is put forward to derive nearly-optimal sensor replacement trajectories for the robotic fleet in a short running time. The composite scheme relies on a swarm of artificial fireflies in which each individual follows the exploratory principles featured by Harmony Search. Infeasible candidate solutions are gradually driven into feasibility under the influence of a weak Pareto dominance relationship. A repair heuristic is finally applied to yield a full-blown solution. To the best of our knowledge, our scheme is the first one in literature that tackles MC2R instances. Empirical results indicate that promising solutions can be achieved in a limited time span.

5.5.2. Energy prediction

One way to improve energy supply for sensor nodes is through ambient energy harvesting from solar, thermal or vibration energy sources coupled with rechargeable energy storage. Wireless sensors have to adapt to the stochastic nature of the energy harvesting sources. We are convinced that predicting the temporal availability of ambient energy resources is vital to plan the harvesting efficiency, optimum resource utilization and energy conservation within sensor nodes. In [20] we propose a novel two stage Autoregressive Weather conditioned Solar Energy Prediction (AWSEP) model which is characterized by low computational complexity and is used to accurately estimate the amount of solar energy that will be harvested in the near future in a particular region. Our algorithm re-learns the model parameters during the prediction processing situations where the prediction error becomes larger than a predefined prediction error threshold mainly because of the unreliable nature of outdoor solar energy sources caused by changes in weather conditions. The proposed AWSEP model performance is evaluated by varying energy harvesting source prediction intervals, sampling rates, trade-offs in prediction accuracy and computational costs using real solar datasets. We concluded that AWSEP algorithm is more accurate, has reduced computational complexity and memory utilization than other prediction schemes in literature. Our proposed algorithm can assist a node to automatically adapt to the changing weather conditions for effective power management and sensing task scheduling.

5.5.3. Servicing sensor nodes

Due to the robots' potential to unleash a wider set of networking means and thus augment the network performance, WSRNs have rapidly become a hot research area. In [5], we elaborate on WSRNs from two unique standpoints: robot task allocation and robot task fulfillment. The former deals with robots cooperatively deciding on the set of tasks to be individually carried out to achieve a desired goal; the latter enables robots to fulfill the assigned tasks through intelligent mobility scheduling.

5.6. Experimenting

Participants: Nathalie Mitton, Julien Vandaele.

One of the goal of the FUN research group is to validate through experimentations and to provide tools for this purpose. Therefore, the FIT platform is deployed, together with a set of tutorials [37]. Nevertheless, we are aware that using testbed platforms for validation is already a great step but it can not satisfy all needs. This is why we also investigate alternatives as emulation. In [28], [32] for instance, we propose a specifically designed experimental setup using a relatively small number of nodes forming a real one-hop neighborhood used to emulate any real WSN. The source node is a fixed sensor, and all other sensors are candidate forwarding neighbors towards a virtual destination. The source node achieves one forwarding step, then the virtual destination position and neighborhood are adjusted. The same source is used again to repeat the process. The main novelty is to spread available nodes regularly following a hexagonal pattern around the central node, used as the source, and selectively use subsets of the surrounding nodes at each step of the routing process to provide the desired density and achieve changes in configurations. Compared to real testbeds, our proposition has the advantages of emulating networks with any desired node distribution and densities, which may not be possible in a small scale implementation, and of unbounded scalability since we can emulate networks with an arbitrary number of nodes. Finally, our approach can emulate networks of various shapes, possibly with holes and obstacles. It can also emulate recovery mode in geographic routing, which appears impossible with any existing approach.

5.7. RFID middleware

Participants: Roberto Quilez, Nathalie Mitton.

The Object Naming System (ONS) is a central lookup service used in the EPCglobal network for retrieving location information about a specific Electronic Product Code (EPC). This centralized solution lacks scalability and fault tolerance and encounters some political issues. In [30], we present the design principles of a fully-distributed multi-root solution for ONS lookup service. In distributed systems, the problem of providing a scalable location service requires a dynamic mechanism to associate identification and location. We design, prototype, and evaluate PRONS, a DHT-based solution for the multi-root problem. We show that PRONS achieves significant performance levels while respecting a number of neutrality requirements.

5.8. VANET

Participants: Enrico Natalizio, Thierry Delot.

Today, thanks to vehicular networks, drivers may receive useful information produced or relayed by neighboring sensors or vehicles (e.g., the location of an available parking space, of a traffic congestion, etc.). In [33], we address the problem of providing assistance to the driver when no recent information has been received on his/her vehicle. Therefore, we present a cooperative scheme to aggregate, store and exchange these events in order to have an history of past events. This scheme is based on a dedicated spatio-temporal aggregation structure using Flajolet-Martin sketches and deployed on each vehicle. Contrary to existing approaches considering data aggregation in vehicular networks, our main goal here is not to save network bandwidth but rather to extract useful knowledge from previous observations. In this paper, we present our aggregation data structure, the associated exchange protocol and a set of experiments showing the effectiveness of our proposal.

In [36], we present a novel vehicular communication protocol, which aims to reduce the effect of broadcast storm problem in VANETs (Vehcular AdHoc NETworks). When the traffic density is above a certain value (e.g., when vehicles are in congested traffic scenarios), one of the most serious problems is the increase of packet collisions and medium contentions among vehicles which attempt to communicate. Our proposed technique, namely Selective Reliable Broadcast protocol (SRB), is intended to limit the number of packet transmissions, by means of opportunistically selecting neighboring nodes, acting as relay nodes. As a result, the number of forwarder vehicles is strongly reduced, while network performance is preserved. SRB belongs to the class of broadcast protocols, and exploits the traditional vehicular partitioning behavior to select forwarders. Each cluster is automatically detected as a zone of interest, whenever a vehicle is approaching, and packets will be forwarded only to selected vehicles, opportunistically elected as cluster-heads. In respect of traditional broadcast approaches, the main strengths of SRB are the efficiency of detecting clusters and selecting forwarders in a fast way, in order to limit the broadcast storm problem. Simulation results have been carried out both in urban and highway scenarios, in order to validate the effectiveness of SRB, in terms of cluster detection and reduction of number of selected forwarders.

MINT Project-Team

6. New Results

6.1. Noisy input filtering for interactive systems

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Noisy signals occur when an original time varying value undergoes undesirable and unpredictable perturbations. These may be caused by things like heat and magnetic fields affecting hardware circuitry, the limits of sensor resolution, or even unstable numerical computation. Noisy signals are a common problem when tracking human motion, particularly with custom sensing hardware and inexpensive input devices like the Kinect or Wiimote.

We developed the 1€ filter ("one Euro filter") is a simple algorithm to filter noisy signals for high precision and responsiveness. It uses a first order low-pass filter with an adaptive cutoff frequency: at low speeds, a low cutoff stabilizes the signal by reducing jitter, but as speed increases, the cutoff is increased to reduce lag. The algorithm is easy to implement, uses very few resources, and with two easily understood parameters, it is easy to tune. When compared with other filters, the 1€ filter shows less lag for a reference amount of jitter reduction [15].

The 1€ filter is already used on a daily basis by many other researchers and companies.

6.2. Transfer functions for subpixel interaction

Participants: Jonathan Aceituno, Géry Casiez [correspondant], Nicolas Roussel.

The current practice of using integer positions for pointing events artificially constrains human precision capabilities (Figure 1). The high sensitivity of current input devices can be harnessed to enable precise direct manipulation "in between" pixels, called subpixel interaction. In [23], we provide a detailed analysis of subpixel theory and implementation, including the critical component of revised control-display gain transfer functions. A prototype implementation is described with several illustrative examples. Guidelines for subpixel domain applicability are provided and an overview of required changes to operating systems and graphical user interface frameworks are discussed.

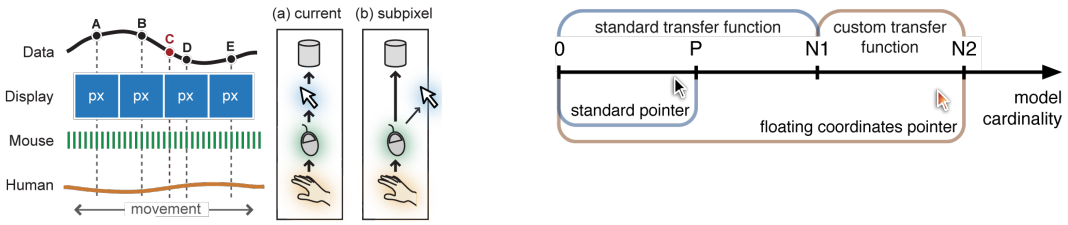
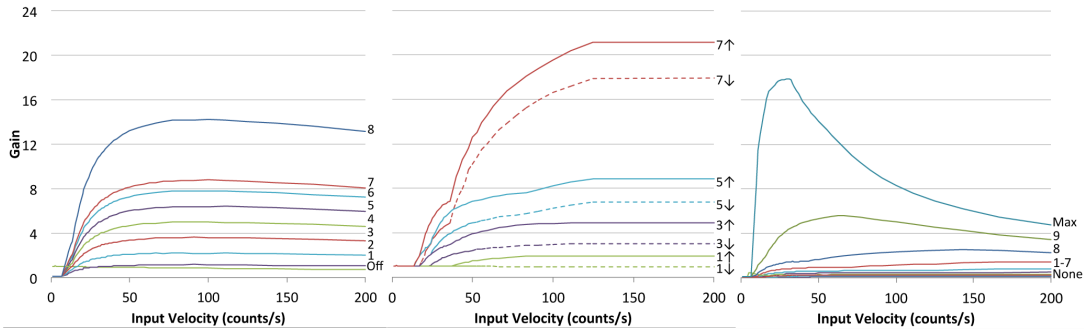


Figure 1. Input mappings: (a) currently, human movements are discretized by mouse sensitivity, then again by display density: data points "in between" pixels like 'C' are unreachable; (b) a subpixel mapping discretizes human movements by mouse sensitivity only, for precise data manipulation (left). Four zones of applicability for subpixel and custom transfer functions (see text for description) (right).

6.3. Transfer functions for scrolling tasks

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Scrolling is controlled through many forms of input devices, such as mouse wheels, trackpad gestures, arrow keys, and joysticks. Performance with these devices can be adjusted by introducing variable transfer functions to alter the range of expressible speed, precision, and sensitivity. However, existing transfer functions are typically "black boxes" bundled into proprietary operating systems and drivers. This presents three problems for researchers: (1) a lack of knowledge about the current state of the field; (2) a difficulty in replicating research that uses scrolling devices; and (3) a potential experimental confound when evaluating scrolling devices and techniques. These three problems are caused by gaps in researchers' knowledge about what device and movement factors are important for scrolling transfer functions, and about how existing devices and drivers use these factors (Figure 2). We fill these knowledge gaps with a framework of transfer function factors for scrolling, and a method for analysing proprietary transfer functions demonstrating how state of the art commercial devices accommodate some of the human control phenomena observed in prior studies [22].



(a) Mac OS X (Generic Device). (b) Microsoft IntelliPoint: up (solid) and down (dashed). (c) Logitech Control Center.

Figure 2. Gain scale factors across input velocity (counts per second) with Mac OS X, Microsoft IntelliPoint (under Windows 7), and Logitech drivers under Mac OS X. Gain is measured as the level of amplification in the system's base unit (pixels per count for Mac OS X and Logitech; lines per count for Microsoft IntelliPoint), and is plotted at varying levels of each driver's respective UI sliders for acceleration.

6.4. Design of transparent tactile stimulators

Participants: Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail [correspondant].

Friction reduction based tactile devices are able to modulate the friction between the fingertip and the active touched surface as a function of fingertip's position. This type of tactile stimulator is thus based on two main components: an active area which vibrates and produces a squeezed air film bearing and a position sensor. Our previous design was made up with a copper plate fully covered by piezo cells, a material which bent when energized by a voltage.

However, this design no longer makes sense when we look forward using tactile feedback on a transparent display. Indeed, for co-localized operation, we can't place piezo cells on the bottom surface of a touch screen since the touched surface would not be transparent; moreover, glass is a non conductive material which complicates the electrical connection.

To cope with these problems, a new design has been introduced. Two copper exciters are firmly bonded on the touch screen to obtain the vibration. These exciters vibrate and propagate their vibration to the glass touch screen. To be efficient, the size of the exciters has to be perfectly adapted to the glass plate. This is why, we not only propose a new way to obtain the vibration of the active area, but we also provide the key design rules of the exciters[19].

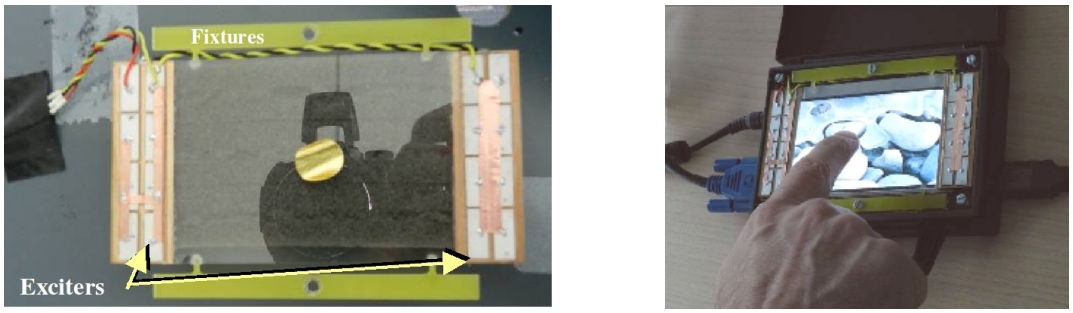


Figure 3. The transparent tactile display, during test procedure (left) and in a co-localized operation (right).

6.5. Methodology for developing textures on friction based interfaces

Participants: Géry Casiez, Thomas Pietrzak, Ludovic Potier, Nicolas Roussel [correspondant], Ibrahim Yapici.

The design of textures for so-called variable friction technologies requires multiple perspectives, which this paper aims to outline and discuss. We first propose a definition of texture and describe the current state of knowledge on their perception. After presenting two technologies for variable friction and comparing them to other tactile interfaces, we describe several particular uses for these devices (Figure 4). We then discuss psychophysical methods for signal perception evaluation and finally discuss methodologies for creating multidimensional tactile content [26].

Figure 4. Examples of textures with increasing complexity in one dimension.

6.6. Hand occlusion on mutitouch surfaces

Participant: Géry Casiez [correspondant].

Operating a computer by directly touching the display surface has many benefits, and in tabletop computing, multi-touch is arguably the most natural form of input. However, with any form of direct input, where the input device and the output display are coincident, the hand and arm cover - or occlude part - of the display. This can be a problem, because compared to manipulating objects on a real tabletop, a tabletop computer is dynamic and can display relevant information, sequential widgets, and system messages in occluded areas.

We examined the shape of hand and forearm occlusion on a multi-touch table for different touch contact types and tasks. Individuals have characteristic occlusion shapes, but with commonalities across tasks, postures, and handedness. Based on this, we create templates for designers to justify occlusion-related decisions and we propose geometric models capturing the shape of occlusion. A model using diffused illumination captures performed well when augmented with a forearm rectangle, as did a modified circle and rectangle model with

ellipse "fingers" suitable when only X-Y contact positions are available (Figure 5). Finally, we describe the corpus of detailed multi-touch input data we generated which is available to the community [24].

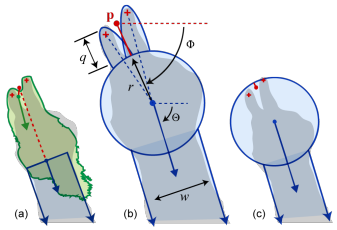


Figure 5. Three occlusion shape models: (a) DI and rectangle; (b) multi-touch circle and rectangle; (c) Vogel et al.

6.7. Indirect multitouch interaction on large screens

Participants: Géry Casiez [correspondant], Jérémie Gilliot, Nicolas Roussel.

Multitouch interaction shows its limits with large display surfaces. Indirect interaction allows to use control surfaces that are much smaller than display surfaces. Absolute indirect interaction raises accuracy problems and relative indirect interaction only allows to interact with a single cursor. We present a relative indirect multitouch interaction technique allowing to create, control, delete several cursors without sacrifying precision for interacting with small objects (Figure 6) [25].

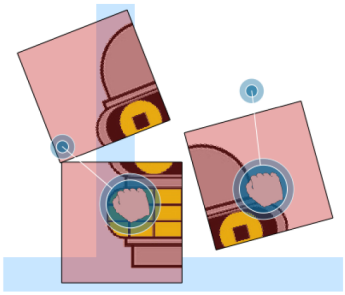


Figure 6. Overview of cursors and cursels used to manipulate two objects.

6.8. Pseudo-rigid movements for flexible multi-finger interactions

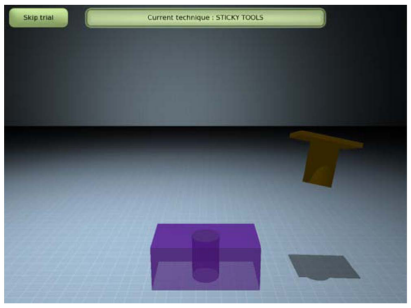
Participants: Laurent Grisoni [correspondant], Yosra Rekik, Nicolas Roussel.

Multi-touch interaction requires a trade-off between users' desires and capabilities and gesture recognition constraints. Current approaches to that problem lack flexibility. The number of fingers used for a gesture usually plays a key part in the recognition process, for example. To increase the flexibility of this process, we proposed the use of *pseudo-rigid movements* [27]. We showed how these movements can be determined in real time from the contact information usually available. We explained how they allow to free the recognition process from the number of fingers used and to move towards multi-movement gestures, independent or coordinated. We also presented an interaction technique that takes advantage of this increased flexibility.

6.9. 3D manipulation on multitouch displays

Participants: Anthony Martinet, Géry Casiez [correspondant], Laurent Grisoni.

Multitouch displays represent a promising technology for the display and manipulation of data. While the manipulation of 2D data has been widely explored, 3D manipulation with multitouch displays remains largely unexplored. Based on an analysis of the integration and separation of degrees of freedom, we propose a taxonomy for 3D manipulation techniques with multitouch displays. Using that taxonomy, we introduce Depth-Separated Screen-Space (DS3), a new 3D manipulation technique based on the separation of translation and rotation. In a controlled experiment, we compared DS3 with Sticky Tools and Screen-Space. Results show that separating the control of translation and rotation significantly affects performance for 3D manipulation, with DS3 performing faster than the two other techniques [13].



		Translation			Rotation		
	Mode	Tx	Ty	Tz	Rx	Ry	Rz
DS3	1*d*	○	○				
	1*d* + 1*i*	○	○	*i*	Z-Technique		
	≥ 2*d*				○	○	○
	≥ 2*d* + 1*i*			*i*	○	○	○

Figure 7. Screen capture of the peg-in-hole task used in the experiment (left). Description of the DS3 technique using the proposed taxonomy (right).

6.10. 3D navigation on multitouch displays

Participants: Clément Moerman, Damien Marchal [correspondant], Nicolas Roussel.

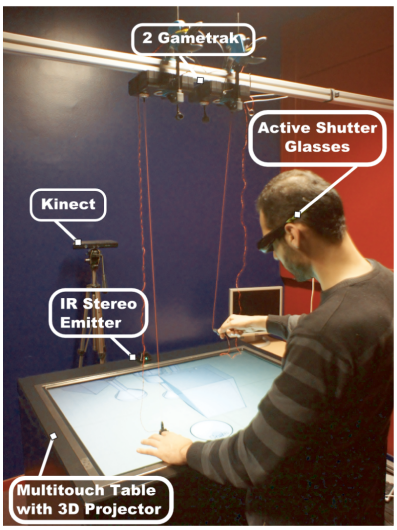
Navigation is one of the elementary tasks of 3d virtual environment. It is composed of two parts: locomotion where there is a physical control of the camera and the wayfinding where a path is found through the environment. Despite being widely studied, there is still need for more efficient and intuitive techniques especially for novice users. Within the context the I-Lab, we worked on a new locomotion technique that combines the advantages of multi-scale navigation and of direct manipulation (Figure 8). The technique, called *Drag'n Go*, was evaluated with a comparative experiment against three other techniques. The results show that *Drag'n Go*: improves performances, reduces learning time and get good user satisfaction either from novice and expert users. The approach and the associated experiment are published in [20].

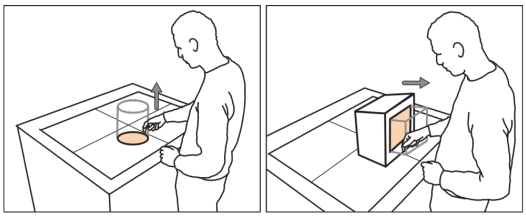
6.11. Modeling on and above a multitouch surface

Participants: Géry Casiez [correspondant], Bruno De Araujo.

We introduced a semi-immersive environment for conceptual design where virtual mockups are obtained from gestures we aim to get closer to the way people conceive, create and manipulate three- dimensional shapes. We presented on-and-above-the-surface interaction techniques following Guiard's asymmetric bimanual model to take advantage of the continuous interaction space for creating and editing 3D models in a stereoscopic environment. To allow for more expressive interactions, our approach continuously combines hand and finger tracking in the space above the table with multi-touch on its surface. This combination brings forth an alternative design environment where users can seamlessly switch between interacting on the surface or in the space above it depending on the task (Figure 9). Our approach integrates continuous space usage with bimanual interaction to provide an expressive set of 3D modeling operations. Preliminary trials with our experimental setup show this as a very promising avenue for further work [17], [16].

Figure 8. With the Drag'n Go method user can navigate in a 3D virtual environment with a multi-touch device. The movement speed is calculated using perspective based progression scale and it is let under the user's control.





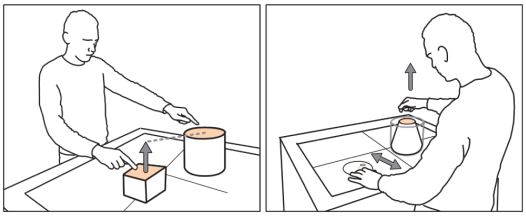


Figure 9. Overview of MockupBuilder setup (left). Examples of face straight extrusion, height constraint and scaling (right).

6.12. Paper-based annotation of digital content from a mobile device

Participant: Thomas Pietrzak [correspondant].

S-Notebook is a hybrid system that makes it possible to take notes on paper about digital content one is exploring on a mobile device (Figure 10). The user can link notes on paper with the content he is currently interacting with, making it possible to reopen it at a later time by tapping the note on his notebook with the digital pen. Therefore he can create bookmarks and hyperlinks on paper notes [21].

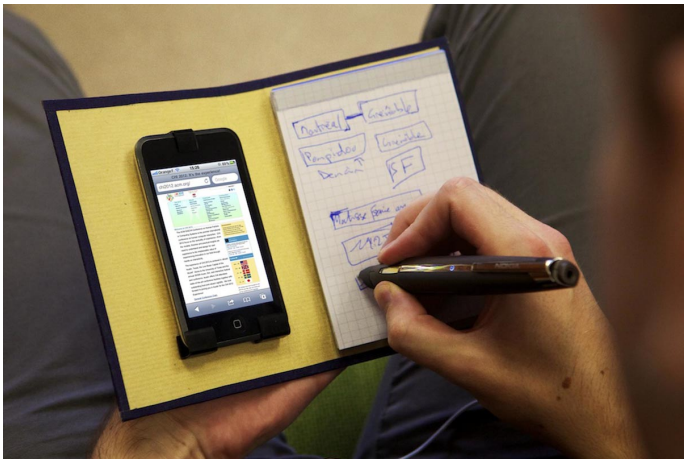


Figure 10. Annotation of digital content on paper.

MODAL Project-Team

6. New Results

6.1. Model for conditionally correlated categorical data

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

An extension of the latent class model is proposed for clustering categorical data by relaxing the classical class conditional independence assumption of variables. In this model, variables are grouped into inter-independent and intra-dependent blocks in order to consider the main intra-class correlations. The dependence between variables grouped into the same block is taken into account by mixing two extreme distributions, which are respectively the independence and the maximum dependence ones. In the conditionally correlated data case, this approach is expected to reduce biases involved by the latent class model and to produce a meaningful model with few additional parameters. The parameters estimation by maximum likelihood is performed by an EM algorithm while a MCMC algorithm avoiding combinatorial problems involved by the block structure search is used for model selection. Applications on sociological and biological data sets bring out the proposed model interest. These results strengthen the idea that the proposed model is meaningful and that biases induced by the conditional independence assumption of the latent class model are reduced. This model was used in September for software components data set of Philippe Merle (ADAM Team Inria Lille).

A conference paper [26] and a poster workshop [35] have been presented. A preprint has been also written [45]. Furthermore, an R package is currently under development.

6.2. Model-based clustering for multivariate partial ranking data

Participants: Christophe Biernacki, Julien Jacques.

The first model-based clustering algorithm dedicated to multivariate partial ranking data has been developed in [43]. This is an extension of the (ISR) model for ranking data published in [4]. The proposed algorithm has allowed to exhibit regional alliances between European countries in the Eurovision contest, which are often suspected but never proved.

6.3. A new probability distribution for ordinal data

Participants: Christophe Biernacki, Julien Jacques.

In [21], a probability distribution for univariate ordinal data is proposed from a stochastic dichotomic search algorithm in a sorting table. Interest of this approach is to give a specific model for ordinal data, without any reference to numerical or nominal data, as it is often the case. The resulting distribution is governed by a position and a dispersion parameter, and is easily estimated by using an EM algorithm.

6.4. Clustering and variable selection in regression

Participants: Christophe Biernacki, Julien Jacques, Loic Yengo.

The works presented in [28] address the issue of simultaneous linear regression and clustering of predictors. A new framework is proposed that both sidesteps optimization challenges and improves prediction performance. In that framework, regression coefficients are assumed to be drawn from a gaussian mixture distribution. Prediction is thus performed using the conditional distribution of the regression coefficients given the data, while clusters are easily derived from posterior distribution in groups given the data.

6.5. Mixture of Gaussians with Missing Data

Participants: Christophe Biernacki, Vincent Vandewalle.

The generative models allow to handle with missing data. This can be easily performed by using the EM algorithm, which has a closed form M-step in the Gaussian setting. This can for instance be useful for distance estimation with missing data. It has been proposed in [18] to improve the distance estimation by fitting a mixture of Gaussian distribution instead of a considering only one Gaussian component. An extension of the previous work including the high setting has been submitted in Neurocomputing journal. This is a joined work with Emil Eirola and Amaury Lendrasse .

A parallel work is in progress on the mixture degeneracy when considering mixture of Gaussians with missing data. It have been experimentally noticed that the degeneracy in this case is particularly slow. This behaviour is different from the usual setting of degeneracy with mixture of Gaussians which is usually rather fast. We are working on the theoretical characterization of this behaviour around a degenerated solution.

6.6. Transfer learning in model-based clustering

Participant: Christophe Biernacki.

In many situations one needs to cluster several datasets, possibly arising from different populations, instead of a single one, into partitions with identical meaning and described by similar features. Such situations involve commonly two kinds of standard clustering processes. The samples are clustered traditionally either as if all units arose from the same distribution, or on the contrary as if the samples came from distinct and unrelated populations. But a third situation should be considered: As the datasets share statistical units of same nature and as they are described by features of same meaning, there may exist some link between the samples. We propose a linear stochastic link between the samples, what can be justified from some simple but realistic assumptions, both in the Gaussian and in the t mixture model-based clustering context [37]. This is a joint work with Alexandre Lourme.

A book chapter about transfer learning (including clustering, classification and regression) has been also published [37]. It is a joint work with Farid Beninel, Charles Bouveyron, Julien Jacques and Alexandre Lourme.

6.7. Gaussian Models Scale Invariant and Stable by Projection

Participant: Christophe Biernacki.

Gaussian mixture model-based clustering is now a standard tool to determine an hypothetical underlying structure into continuous data. However many usual parsimonious models, despite their appealing geometrical interpretation, suffer from major drawbacks as scale dependence or unsustainability of the constraints by projection. In this work we present a new family of parsimonious Gaussian models based on a variance-correlation decomposition of the covariance matrices. These new models are stable by projection into the canonical planes and, so, faithfully representable in low dimension. They are also stable by modification of the measurement units of the data and such a modification does not change the model selection based on likelihood criteria. We highlight all these stability properties by a specific geometrical representation of each model. A detailed GEM algorithm is also provided for every model inference. Then, on biological and geological data, we compare our stable models to standard geometrical ones.

This work is was presented as a poster to workshop [31] and is also a preprint [41] currently in revision in an international journal. This is a joint work with Alexandre Lourme.

6.8. Decorrelating variables in high dimension for linear regression

Participants: Christophe Biernacki, Clément Thery.

Databases from the steel industry are often large (very long process with many parameters) and have strong correlations between variables. Some variables may be written directly in terms of other via physical models or related by definition. Moreover the process, which is specific to the type of finished product, conditions most of the process parameters and therefore induces strong correlations between variables. The main idea is to consider some form of sub-regressions models, some variables defining others. We can then remove temporarily some of the variables to overcome ill-conditioned matrices inherent in linear regression and then reinject the deleted information, based on the struc- ture that links the variables. The final model therefore takes into account all the variables but without suffering from the consequences of correlations between variables or high dimension. This research is placed in a steel industry context (Arcelor-Mittal Dunkerque).

The work has been presented to a conference [27] and as a poster to a workshop [36]. It is a joint work with Gaétan Loridant from Arcelor-Mittal.

6.9. Model-based clustering for multivariate functional data

Participants: Julien Jacques, Cristian Preda.

We developed in [19] an extension of the model-based clustering algorithm for univariate functional data proposed in [20], [23], [11] to the case of multivariate functional data. For this, multivariate functional principal components analysis is defined and a parametric mixture model is proposed and estimated by an EM-like algorithm. Results on simulated and real datasets have shown the efficiency of the proposed method.

6.10. A method to combine combinatorial optimization and statistics to mine high-throughput genotyping data

Participants: Julie Hamon, Julien Jacques, Clarisse Dhaenens.

In the context of genomic analysis (collaboration with Genes Diffusion), dealing with high-throughput genotyping data, the objective of our study is to select a subset of SNPs (single nucleotide polymorphisms) explaining a trait of interest. We propose in [33] and [32] a method combining combinatorial optimization and statistics to extract a subset of interesting SNPs. The combinatorial part aims at exploring in a efficient way the large search space induced by the large number of possible subsets and statistics are used to evaluate the selection. We propose a first method based on an ILS (iterated local search) and using a regression. Three criteria used to evaluate the quality of the regression are compared. One of them (the k-fold validation) shows better performance. We also compare this approach to classical statistical approaches on simulated datasets. Results are promising as the proposed approach outperforms most of these statistical approaches.

6.11. Wavelet based clustering using mixed effects functional models

Participant: Guillemette Marot.

Curve clustering in the presence of inter-individual variability has longly been studied, especially using splines to account for functional random effects. However splines are not appropriate when dealing with high-dimensional data and can not be used to model irregular curves such as peak-like data. We propose a wavelet based clustering procedure ([6]) and apply it to high dimensional data. We suggest a dimension reduction step based on wavelet thresholding adapted to multiple curves and using an appropriate structure for the random effect variance, we ensure that both fixed and random effects lie in the same functional space even when dealing with irregular functions that belong to Besov spaces. In the wavelet domain, our model resumes to a linear mixed-effects model that can be used for a model-based clustering algorithm and for which we develop an EM- algorithm for maximum likelihood estimation. An R package curvclust implementing this procedure has been posted this year to the CRAN, the official website of the R software.

6.12. Comparison of normalisation procedures in RNA-sequencing before differential analysis

Participant: Guillemette Marot.

The continuing technical improvements and decreasing cost of next-generation sequencing technologies have made RNA sequencing (RNA-seq) a popular choice for gene expression studies. Several methods for the normalization of RNAseq data (removal of errors due to the small number of samples, corrections for sequence composition) have been proposed in recent years. With the Statomique Consortium, we have compared seven normalisation methods, discarded two out of them (although still widely used). We give practical recommendations on the appropriate normalization method to be used and its impact on the differential analysis of RNA-seq data in the paper ([14]).

6.13. Change point detection algorithm

Participant: Alain Célisse.

We develop a new change-point detection algorithm where focus is given to detect changes in the whole distribution of data. This challenging problem is addressed by use of kernels which enable us to deal with non-vectorial data of aby type (graphs, DNA sequences, etc). A preprint has been submitted ([46]).

6.14. Cross validation algorithms

Participant: Alain Célisse.

The performance of Cross-validation (CV) algorithms are assessed for estimating the risk as well as for model selection. Whereas optimality of leave-one-out (LOO) cross-validation is proved for risk estimation, it is no longer the case for model selection. In the latter setup, conditions are derived that lead to optimality for leave-p-cross-validation (LPO) when p is larger than 1. See for details [47].

6.15. Stochastic Block Model

Participant: Alain Célisse.

The convergence of maximum likelihood and variational estimators in a random graph model called Stochastic Block model is addressed. To the best of our knowledge, these are the first results providing consistency for maximum likelihood and variational estimators in that model. See [5].

6.16. Approximations for scan statistics.

Participants: Alexandru Amarioarei, Cristian Preda.

Accurate approximations for the distribution of extremes of 1-dependent stationary sequences are developed (see [38]). Viewed as maximum of some particular sequence of 1-dependent random variables, we provide sharp error bounds and approximations for the distribution of the three-dimensional scan statistics (see [39]). The Binomial and Poisson models are considered.

MOSTRARE Project-Team

6. New Results

6.1. Modeling XML document transformations

Participants: Joachim Niehren, Angela Bonifati, Sophie Tison, Sławek Staworko, Aurélien Lemay, Anne-Cécile Caron, Yves Roos, Benoît Groz, Antoine Ndione, Tom Sebastian.

XML Schema Validation Groz, Staworko et. al. [26] present a new algorithm that tests determinism of regular expressions in linear time. All regular expressions used in DTDs and XML Schemas are required to be deterministic by the recommendation of the W3C. Whether this is the case can indeed been tested in linear time, as shown in this paper. The best known previous algorithm, which was based on the Glushkov automaton, required O(σ|e|) time, where σ is the number of distinct symbols in e. They also show that matching a word w against a deterministic regular expression e can be achieved in combined linear time O(|e| + |w|) for a wide range of cases.

Staworko et. al. studied bounded repairability for regular tree languages modulo the tree edit distance [28].

Ndione, Niehren, and Lemay [33] present a new probabilistic algorithm for approximate membership of words to regular languages modulo the edit distance on words. In the context of XML, this algorithm is relevant for sublinear DTD validity testing. The time complexity of the algorithm is independent of the size of the input word and polynomial in the size of the input automaton and the inverse error precision. All previous property testing algorithms for regular languages run in exponential time.

XML Query Answering Debarbieux, Niehren, Sebastian et. al. [32] present new algorithms for early XPath node selection on XML Streams. Early selection and rejection is crucial for efficiency, while earliest selection and rejection has high computational complexity in the general case. In contrast to all previous approaches, there algorithm does not rely on any expensive static analysis method. Instead, it is based on a compiler from XPath to nested word automata with selection and rejection states that they introduce. They cover a large fragment of downward XPath, with the main restriction that negation is forbidden above descendant axis and disjunctions. Non-determinism is used to deal with descendant axis and disjuctions. High run-time efficiency in practice is obtained by on-the-fly determinization for nested word automata, even in cases where static determinization produces automata of more than exponential size. Our experimental results confirm a very high efficiency in space and time. An implementation of our FXP/QuiXPath system is freely available and used for industrial transfer in the QuiXProc system.

Staworko et. al. tackled prioritized repairing and consistent query answering in relational databases in [20].

External Cooperations with other teams in Lille lead to the following publications [19], [31], [30].

6.2. Machine learning for XML document transformations

Participants: Adrien Boiret, Jean Decoster, Pascal Denis, Jean-Baptiste Faddoul, Antonino Freno, Gemma Garriga, Rémi Gilleron, Mikaela Keller, Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Sławek Staworko, Marc Tommasi, Fabien Torre.

Learning XML Queries. Staworko et. al. [29] proposed learning twig and path queries.

Niehren, Champavère, Gilleron, and Lemay [34] propose new algorithm and learnability result for XML query induction based on schema-guided pruning strategies. Pruning strategies impose additional assumptions on node selection queries that are needed to compensate for small numbers of annotated examples. The class of regular queries that are stable under a given schema-guided pruning strategy was distinguished and shown to be learnable with polynomial time and data. The learning algorithm is obtained by adding pruning heuristics to the traditional learning algorithm for tree automata from positive and negative examples. While justified by a formal learning model, their learning algorithm for stable queries also performs very well in practice of XML information extraction.

Learning XML Transformations. Boiret, Lemay, and Niehren [21] solved the long open question of how to learn rational functions with polynomial time and data. Rational functions are transformations from words to words that can be defined by deterministic string transducers with lookahead. No previous learning results for classes of transducers with look-ahead existed, so this results is relevant for learning XML transformations defined by transducers with look-ahead, as with XSLT.

Multi-task Learning. We address the problem of multi-task learning with no label correspondence among tasks. In [22], Faddoul, Chidlovskii, Gilleron and Torre propose the multi-task Adaboost algorithm with Multi-Task Decision Trees as weak classifiers. They conduct experiments on multi-task datasets, including the Enron email set and Spam Filtering collection. Faddoul successfully defended his PhD thesis [16] in June 2012.

Probabilistic models for large graphs. We propose new approaches for the statistical analysis of large-scale undirected graphs. The guiding idea is to exploit the spectral decomposition of subgraph samples, and in particular their Fiedler eigenvalues, as basic features for density estimation and probabilistic inference. In [24], Freno, Keller, Garriga, and Tommasi develop a conditional random graph model for learning to predict links in information networks (such as scientific coauthorship and email communication). In [25], Freno, Keller, and Tommasi propose instead to estimate joint probability distributions through (non-linear) random fields, applying the resulting model to graph generation and link prediction.

Learning in Multiple graphs Ricatte, Garriga, Gilleron and Tommasi focus on learning from several sources of heterogeneous data. They represent each source as a graph of data and they propose to combine the multiple graphs with the help of small number of labeled nodes. They obtain a kernel that can be used as input to different graph-learning tasks such as node classification and clustering. The paper is under submission. Along a collaboration with physicians, Keller and Tommasi consider graphs that represents the structural connectivity of the brain (connectome). They develop a spatially constrained clustering method, combining heterogeous descriptions of the same objects through the graph of neighborhood on the cortex and the graph of connectivity. The paper is under submission.

Starting PhDs Boneva, Bonifati and Staworko started to supervise the PhD of R. Ciucanu on learning cross-model database mappings. Denis and Tommasi has begun to supervise the PhD of David Chatel on guided clustering for graphs (of texts).

NON-A Project-Team

5. New Results

5.1. Model-free control

Participants: Cédric Join, Samer Riachy.

The achievements obtained in 2012 are as follows:

- The model-free control approach is applied to a complex nonlinear model describing the dynamics of a traffic flow in [24]. The robustness with respect to external disturbances is shown by numerical simulations.
- Model-free control is applied to a magnetic bearing in [56], which is a quite important industrial device. The experimental results are compared to those obtained via other control techniques.
- "Model-free" control and the related "intelligent" proportional-integral controllers are successfully applied to freeway ramp metering control in [47]. Implementing the proposed control strategy is straightforward. Numerical simulations need the identification of quite complex quantities like the free flow speed and the critical density. This is achieved due to new estimation techniques, where the differentiation of noisy signals plays a key role.

5.2. Algebraic technique for estimation, differentiation and its applications

Participants: Cédric Join, Mamadou Mboup, Wilfrid Perruquetti, Rosane Ushirobira, Olivier Gibaru.

Elementary techniques from operational calculus, differential algebra, and noncommutative algebra lead to a new algebraic approach for estimation and detection. It is investigated in various areas of applied sciences and engineering. The following lists only some applications:

- The paper [30] proposes an algebraic method to fault diagnosis for uncertain linear systems. The main advantage of this new approach is to realize fault diagnosis only from knowledge of input and output measurements without identifying explicitly model parameters. Using tools and results of algebraic identification and pseudospectra analysis, the issues of robustness of the proposed approach compared to the model order and noise measurement are examined.
- The aim of [79], [84] is to develop an algebraic approach to estimate human posture in the sagittal plane using inertial measurement unit providing accelerations and angular velocities. For this purpose the issue of the estimation of the amplitude, frequency and phase is addressed for a biased and noisy sum of three sinusoidal waveform signals on a moving time horizon. Since the length of the time window is small, the estimation must be done within a fraction of the signal's period. The problem is solved via algebraic techniques.
- An application of algebraic estimation approach for estimation of option pricing and dynamic hedging is given in [66].
- A model-based online fault-diagnosis scheme for an electromagnetically supported plate is presented in [73] as an example of a nonlinear and open-loop unstable system. First, residuals for sensor as well as for actuator faults are generated using algebraic derivative estimators. Then, the robust detection and isolation of step-like sensor and actuator faults is presented.
- The paper [57] uses the extreme value theory for threshold selection in a previously proposed algebraic spike detection method. The algebraic method characterizes the occurrence of a spike by an irregularity in the neural signal and devises a nonlinear (Volterra) filter which enhances the presence of such irregularities.

- The papers [39], [40] generalize the algebraic method from the integer order to the fractional order for estimating the fractional order derivatives of noisy signals. The proposed fractional order differentiator is deduced from the Jacobi orthogonal polynomial filter and the Riemann-Liouville fractional order derivative definition. Exact and simple formula for this differentiator is given where an integral formula involving Jacobi polynomials and the noisy signal is used without complex mathematical deduction. Hence, it can be used both for continuous-time and discrete-time models. The comparison between our differentiator and the recently introduced digital fractional order Savitzky-Golay differentiator is given in numerical simulations so as to show its accuracy and robustness with respect to corrupting noises.

5.3. Observability and observer design for nonlinear systems

Participants: Jean-Pierre Barbot, Wilfrid Perruquetti, Gang Zheng, Denis Efimov.

Observability analysis and observer design are important issues in the field of control theory. Some recent results are listed below:

- The problem of observer design for fault detection in a class of nonlinear systems subject to parametric and signal uncertainties is studied in [22]. The design procedure includes formalized optimization of observer free parameters in terms of trade-offs for fault detection performance and robustness to external disturbances and model uncertainties. The technique makes use of some monotonicity conditions imposed on the estimation error dynamics. Efficiency of the proposed approach is demonstrated through the Oscillatory Failure Case in aircraft control surface servo-loops.
- An algorithm for the frequency and bias identification of a harmonic signal is presented in [14], [15]. The solution is based on an adaptive observer technique and the hybrid systems method.
- An influence of a singular manifold of non observable states on reconstruction of chaotic attractors is analysed in [25]. The probability of visits of the observability singularity manifold and the relative time spent in the observability singularity manifold are introduced.
- In [36], the cluster structured sparse signals are investigated. Under the framework of Bayesian compressive sensing, a hierarchical Bayesian model is employed to model both the sparse prior and cluster prior, then Markov Chain Monte Carlo (MCMC) sampling is implemented for the inference. Unlike the state-of-the-art algorithms, which are also taking into account the cluster prior, the proposed in [36] algorithm solves the inverse problem automatically–prior information on the number of clusters and the size of each cluster is unknown.
- The papers [54], [86], [87] present a new approach for observer design for a class of nonlinear singular systems which can be transformed into a special normal form. The interest of the proposed form is to facilitate the observer synthesis for the studied nonlinear singular systems. Necessary and sufficient geometrical conditions are deduced in order to guarantee the existence of a diffeomorphism, which transforms the studied nonlinear singular systems into the proposed normal form.
- The paper [85] investigates the observer design problem of for linear switched system with disturbance jumps. Detection of active sub-system and finite time estimation of states are respectively discussed. A switched finite time observer is proposed to guarantee the finite time convergence independent of the disturbance jumps.
- The paper [71], [72] proposes a new observer scheme for chaotic and hyperchaotic systems. Firstly, a classical impulsive observer is investigated for Lorenz chaotic system. This approach is based on sufficient conditions for stability of impulsive dynamical systems. After, an hybrid observer is proposed for hypoerchaotic systems. In the paper [70], a new method of strange attractor identification, under sparse measurement, is proposed this method is based on the concept of compressive sensing. For this, some particular impulsive observers have been presented with a decision scheme linked to diagnosis method, the identification of the strange attractor and state observation are done.

- The problem of state reconstruction for nonlinear differential-algebraic systems with unknown inputs is studied in [51].
- In the paper [26] the design of observers for nonlinear systems with unknown, time-varying, bounded delays, on both state and input for a class of nonlinear systems is proposed. Furthermore, the feasibility of the proposed strategy is illustrated by a numerical example.

5.4. Sliding mode control estimation

Participants: Jean-Pierre Barbot, Wilfrid Perruquetti, Denis Efimov, Thierry Floquet.

Sliding mode algorithms are very popular for finite-time estimation and regulation. The recent results obtained by the group are as follows:

- The issues of a higher order sliding mode controller realization under actuator saturation and quantization have been analysed in [37]. The zig-zag solutions are introduced and analysed.
- The problem of design of interval observers for linear-parameter-varying systems, containing non detectable or non strongly observable parts, is addressed in [18], [63], [62] applying the higher order sliding mode algorithms. Application of sliding mode observers leads to accuracy improvement in the system.
- In [32] an anomaly signal detection in communication networks is studied by control theory techniques. Several classes of sliding mode observers are proposed for a fluid flow model of the transmission control protocol (TCP)/internet protocol network. Comparative simulations via network simulator NS-2 show the enhancement brought by a higher order sliding mode observer. The efficiency of this observer opens the way toward observing traffics with real TCP flow characteristics.
- In [80], [42], [41] the problem of continuous and discrete state estimation for a class of linear switched systems is studied. The class of systems under study can contain non-minimum phase zeros in some of their "operating modes". The conditions for exact reconstruction of the discrete state are given using structural properties of the switched system. The state-space is decomposed into the strongly observable part, the nonstrongly observable part and the unobservable part, to analyze the effect of the unknown inputs. A state observer based on high-order sliding-mode and Luenberger-like observers is proposed. For the case when the exact reconstruction of the state cannot be achieved, the ultimate bounds on the estimation errors are provided. In [41] this technique has been applied to fault detection in switched systems.
- The paper [55] aims, firstly to highlight the possibility of recovering a message included in a chaotic continuous time delay system, secondly to show that it is possible to use the third order sliding mode in order to recover directly all the states and the unknown input (message), thirdly to illustrate the robustness of the proposed observer with respect to a noisy signal. This work is based on the concept of left invertibility and recent advances in sliding mode observers.
- The problem of estimation of discrete and continuous states for switched systems applying higher order sliding mode observers and projection is investigated in the papers [68], [67].

5.5. Non-linear, Sampled and Time-delay systems

Participants: Jean-Pierre Richard, Lotfi Belkoura, Gang Zheng, Denis Efimov, Wilfrid Perruquetti.

Nonlinearities, sampling, quantization and time-delays cause serious obstructions for control and observer design in many fields of techniques and engineering (e.g. networked and internet systems, distributed systems etc.). The proposed by the team algebraic approach suits well for estimation and regulation in such a type of systems. The recent results are listed below:

- A new type of stability is introduced and its equivalent Lyapunov characterization is presented in [16]. The problem of global stability for the compact set composed of all invariant solutions of a nonlinear system (several equilibriums, for instance) is studied. It is shown that several well-known multi-stable systems satisfy this new stability property.

- A new state-dependent sampling control is proposed in [23], [65], which enlarges the sampling intervals of state feedback control. The case of linear time invariant systems with time delays is considered that guarantees the exponential stability of the system origin for a chosen decay rate. The approach is based on LMIs obtained from the sufficient Lyapunov-Razumikhin stability conditions.
- Nonlinear feedback design for fixed-time stabilization of linear control systems is studied in [31]. Nonlinear control algorithms of two types are presented for uncertain linear plants. Controllers of the first type are stabilizing polynomial feedbacks that allow to adjust a guaranteed convergence time of system trajectories into selected neighborhood of the origin independently on initial conditions. Controllers of the second type are modifications of the second order sliding mode control algorithms. They provide global finite-time stability of the closed-loop system and allow to adjust a guaranteed settling time independently on initial conditions. Control algorithms are presented for both single-input and multi-input systems.
- The problem of natural wave control is addressed in [17], which involves steering a lattice of oscillators towards a desired natural (i.e. zero-input) assignment of energy and phase across the lattice. This problem is formulated and solved for lattices of linear oscillators via a passivity-based approach.
- The verification problems for transition systems enriched with a metric structure is analysed in [27]. The main novelty compared to an algorithm presented recently by Lerda et al. [2008] consists in introducing a tuning parameter, which improves the performance drastically. A procedure that allows one to prove unbounded safety from the result of the bounded safety algorithm via a refinement step is also established. The algorithm to handle bounded liveness verification is adapted.
- The problem of finite-time output stabilization of the double integrator is addressed in [52] applying the homogeneity approach. A homogeneous controller and a homogeneous observer are designed (for different degree of homogeneity) ensuring the finite-time stabilization. Their combination under mild conditions is shown to stay homogeneous and finite-time stable as well.
- The notes [76], [77] are dedicated to the stability analysis of bilinear sampled-data systems, controlled via a linear state feedback static controller. A zero order hold device is used. The purpose is to find a constructive way to calculate the maximum allowable sampling period (MASP) that guarantees the local stability of the system. The proposed stability conditions are formulated as linear matrix inequalities (LMI).
- The works [75], [74] concern the adaptation of sampling times for linear time invariant systems controlled by state feedback. Complementary to various works that guarantee stabilization independently of changes in the sampling rate, there the conditions to design stabilizing sequences of sampling instants is provided. In order to reduce the number of these sampling instants, a dynamic scheduling algorithm optimizes, over a given sampling horizon, a sampling sequence depending on the system state value. The proofs are inspired on switched system techniques combining Lyapunov functions and LMI optimization.
- The mechanism of entrainment to natural oscillations in a class of (bio)mechanical systems described by linear models is investigated in [61]. A nonlinear control strategy (based on the speed gradient control algorithm) is analyzed providing the system oscillation in resonance mode with a natural frequency. It ensures an energy-optimal entrainment performance robustly against perturbations in system parameters in a finite time.
- The paper [29] considers a networked control loop, where the plant is a "slave" part, and the remote controller and observer constitute the "master". Since the performance of Networked Control Systems (NCS) depends on the Quality of Service (QoS) available from the network, a controller is designed that takes into account qualitative information on the QoS in real time.
- In the paper [50], the theory of non-commutative rings allows determining whether or not there exists an equation called algebraically essential in order to estimate the delay on a nonlinear system. From this equation, it is shown that this equation is generally not enough to guarantee the delay estimation, thus the notion of persistent signal with respect to delay estimation is introduced.

Furthermore, based on the definitions of algebraically essential equation and of persistent signal, a delay estimation algorithm is proposed. Some simulation results have been presented in order to highlight the robustness (with respect to measurement noise) of the proposed algorithm.

- The problem of algebraic identifiability for linear and nonlinear dynamical systems is considered in [88].

5.6. Interval control and estimation

Participants: Denis Efimov, Wilfrid Perruquetti.

In many cases due to parametric and/or signal uncertainties presented in a plant model it is not possible to design a conventional observer, which provides a point-wise estimate of state in a finite time or asymptotically. In this case it is still frequently possible to apply the interval observer techniques, which generate an estimate on the interval of the admissible values of the state at the current instant of time. The recent results are listed below:

- The problem of output stabilization of a class of nonlinear systems subject to parametric and signal uncertainties is studied in [20], [21]. First, an interval observer is designed estimating the set of admissible values for the state. Next, it is proposed to design a control algorithm for the interval observer providing convergence of interval variables to zero, that implies a similar convergence of the state for the original nonlinear system. An application of the proposed technique shows that a robust stabilization can be performed for linear time-varying and linear-parameter-varying systems without assumption that the vector of scheduling parameters is available for measurements.
- The problem of interval observer design for a class of observable nonlinear systems is studied in [33]. It is shown that under some mild conditions a Hurwitz matrix can be transformed to a Hurwitz and Metzler one using a real similarity transformation.
- The work [64] is devoted to interval observer design for Linear Time Varying (LTV) systems and a class of nonlinear time-varying systems in the output canonical form. An interval observer design is feasible if it is possible to calculate the observer gains making the estimation error dynamics cooperative and stable. It has been shown that under some mild conditions the cooperativity of an LTV system can be ensured by a static linear transformation of coordinates. The case of a time-varying transformation for periodic systems is considered in the work [64].
- The problem of actuator fault detection for flat systems using the sliding-mode differentiation and the interval constraint satisfaction technique has been analysed in [43].

5.7. Applications

Participants: Jean-Pierre Richard, Jean-Pierre Barbot, Mamadou Mboup, Gang Zheng, Denis Efimov, Wilfrid Perruquetti, Olivier Gibaru, Samer Riachy.

As it was mentioned, Non-A is a kind of "method-driven" project, which deals with different aspects of finite-time estimation and control. Thus different applications are possible, ones touched this year are as follows:

- The global stabilization of a ball & beam through saturated control, which imposes restrictions on the reactivity of the closed loop, is studied in [91], [81]. A modified design for the classical ball & beam system is presented. The beam is driven by two actuators. In comparison to the classical system, this design offers an additional degree of freedom, which is the vertical motion of the beam. We show that the new design offers the possibility to get rid of the closed loop low reactivity restriction. Two nonlinear controllers to steer the trajectories of the system towards a final desired position are proposed.
- In papers [48], [49] a new class of power converters is studied (Parallel Multicell Chopper). The topology of these chopper is based on a combination of n switching cells interconnected via independent inductors. This type of choppers is a new DC/DC static power converter which has an output current equals to n times the source current where n is the number of cells . After recalling the dynamical equations of the converter, its hybrid dynamical behavior and properties are highlighted. This particular hybrid system induces new and difficult control problems, such problem can be tackled by a new control concept based on Petri net.

- The paper [69] addresses the problem of power management of a hydrogen fuel cell system combined with super capacitors under high load variations in an electric vehicle. The singular perturbation theory is used for the control and coordination of two converters. The Lyapunov theory is used for analysis.
- Combined feedforward/feedback control algorithm for highly nonlinear systems was proposed on the basis of the approximating hybrid model in [28]. The designed MIMO controller enables simultaneous control of the air-to-fuel ratio and torque for injector automobile engines. The theoretical results were validated experimentally with physical cars.
- A spike sorting method for multi-channel recordings is proposed in [35]. The proposed method uses an iterative application of Independent Component Analysis (ICA) algorithm and a deflation technique in two nested loops. The results suggest that the proposed solution significantly improves the performance of ICA in spike sorting.
- In the paper [83] an algorithm for a particular change-point detection problem is proposed, where the frequency band of the signal changes at some points in the time axis. Apart from detecting the change-points, the proposed algorithm is also able to estimate the frequency bands. The main idea of the algorithm is to consider a simple local bandlimited model to represent the input signal in each sliding time window.
- The papers [60], [58], [59] present a new sensorless parameter identification method for permanent magnet stepper motors. Current sensors are assumed available, but position and velocity sensors are not. Data is obtained with open-loop voltage commands at multiple speeds. A new reference frame is proposed that presents advantages similar to the standard $d-q$ frame, but without the need for a position sensor. The method exploits carefully derived linear parameterizations and a least-squares algorithm. In one case, overparameterization is resolved using elimination theory. Overall, the parameters identified using the new procedure are found to be very close to those obtained with position sensors. The approach is potentially applicable to other types of synchronous motors as well.
- In the paper [78], an improvement of the dynamic accuracy of a flexible robot joint is addressed. Based on the observation of the measured axis deformation, a simplified elastic joint model is deduced. In the first step, the non-linear model component's is analyzed and identified in the cases of the gravity bias and the friction term. In the second step, a non asymptotically algebraic fast identification of the oscillatory behavior of the robot axis is introduced. Finally, the performances of the identification approach are exploited in order to improve the dynamic accuracy of a flexible robot axis. This is done experimentally by the combination of the adaptation of the jerk time profile to reduce the end-point vibration and the model-based precompensation of the end-point tracking error.
- Localizability of unicycle mobiles robots is analysed in [82] from an algebraic point of view. A sensibility study leads to a new fusion algorithm in the multi landmark case using as a basis the posture differentiation based estimator.
- The problem of early detection of oscillatory failures for aircrafts is addressed in [38]. The proposed solution is based on a finite-time sliding-mode differentiator and a hybrid optimization scheme.
- The H_∞ control design under time-varying delays and uncertainties, which ensures the stability and performance (synchronization/transparency) between the master and slave manipulators, is proposed in [46], [44], [45]. The design of the controller based on a proposed control scheme, which is performed by using LMI optimization based on Lyapunov-Krasovskii functionals and H_∞ control theory.

RMOD Project-Team

5. New Results

5.1. Object serializer

Participants: Martin Dias [Correspondant], Mariano Martinez-Peck, Stéphane Ducasse.

Fuel: A Fast General Purpose Object Graph Serializer Since objects need to be stored and reloaded on different environments, serializing object graphs is a very important activity. There is a plethora of serialization frameworks with different requirements and design trade-offs. Most of them are based on recursive parsing of the object graphs, an approach which often is too slow. In addition, most of them prioritize a language-agnostic format instead of speed and language-specific object serialization. For the same reason, such serializers usually do not support features like class-shape changes, global references or executing pre and post load actions. Looking for speed, some frameworks are partially implemented at Virtual Machine (VM) level, hampering code portability and making them difficult to understand, maintain and extend. That is why we work on Fuel, a general-purpose object serializer based on these principles: (1) speed, through a compact binary format and a pickling algorithm which invests time in serialization for obtaining the best performance on materialization; (2) good object-oriented design, without special help at VM; (3) serialize any object, thus have a full-featured language-specific format. We implement and validate this approach in Pharo, where we demonstrate that Fuel is faster than other serializers, even those with special VM support. The extensibility of Fuel made possible to successfully serialize various objects: classes in Newspeak, debugger stacks, and full CMS object graphs [11].

5.2. Cycles and dependencies

Participants: Stéphane Ducasse [Correspondant], Nicolas Anquetil, Muhammad Bhatti.

OZONE: Layer Identification in the presence of Cyclic Dependencies A layered software architecture helps understanding the role of software entities (e.g., packages or classes) in a system and hence, the impact of changes on these entities. However, the computation of an optimal layered organization in the presence of cyclic dependencies is difficult. We develop an approach that (i) provides a strategy supporting the automated detection of cyclic dependencies, (ii) proposes heuristics to break cyclic dependencies, and (iii) computes an organization of software entities in multiple layers even in presence of cyclic dependencies. Our approach performs better than the other existing approaches in terms of accuracy and interactivity, it supports human inputs and constraints. We compare this approach to existing solutions and apply it on two large software systems to identify package layers. The results are manually validated by software engineers of the two systems [12].

Efficient Retrieval and Ranking of Undesired Package Cycles in Large Software Systems Many design guidelines state that a software system architecture should avoid cycles between its packages. Yet such cycles appear again and again in many programs. We believe that the existing approaches for cycle detection are too coarse to assist developers to remove cycles from their programs. We design an efficient algorithm that performs a fine-grained analysis of cycles among application packages. In addition, we define multiple metrics to rank cycles by their level of undesirability, prioritizing cycles that are the more undesired by developers. We compare these multiple ranking metrics on four large and mature software systems in Java and Smalltalk [14].

Resolving cyclic dependencies between packages with Enriched Dependency Structural Matrix Dependency Structural Matrix (DSM) is an approach originally developed for process optimization. It has been successfully applied to identify software dependencies among packages and subsystems. A number of algorithms have been proposed to compute the matrix so that it highlights patterns and problematic dependencies between subsystems. However, existing DSM implementations often miss important information to fully support reengineering effort. For example, they do not clearly qualify and quantify problematic relationships, information that is crucial to support remediation tasks. We propose Enriched Dependency Structural Matrix (eDSM), which provides small multiple views and micro-macro readings by adding fine-grained information in each cell of the matrix. Each cell is enriched with contextual information about (i) the type of dependencies (inheritance, class reference, etc.), (ii) the proportion of referencing entities, (iii) the proportion of referenced entities. We distinguish independent cycles and stress potentially simple fixes for cycles using coloring information. This work is language independent and has been implemented on top of the Moose software analysis platform.We improved the cell content information view based on user feedback and performed multiple validations: two different case studies on Moose and Seaside software; one user study for validating eDSM as a usable approach for developers. Solutions to problems identified with eDSM have been performed and retrofitted in analyzed software [13].

5.3. Warnings and bugs

Participants: Simon Allier [Correspondant], Andre Hora, Nicolas Anquetil, Muhammad Bhatti, Stéphane Ducasse.

A Framework to Compare Alert Ranking Algorithms To improve software quality, rule checkers statically check if a software contains violations of good programming practices. On a real sized system, the alerts (rule violations detected by the tool) may be numbered by the thousands. Unfortunately, these tools generate a high proportion of "false alerts", which in the context of a specific software, should not be fixed. Huge numbers of false alerts may render impossible the finding and correction of "true alerts" and dissuade developers from using these tools. In order to overcome this problem, the literature provides different ranking methods that aim at computing the probability of an alert being a "true one". We propose a framework for comparing these ranking algorithms and identify the best approach to rank alerts. We have selected six algorithms described in literature. For comparison, we use a benchmark covering two programming languages (Java and Smalltalk) and three rule checkers (FindBug, PMD, SmallLint). Results show that the best ranking methods are based on the history of past alerts and their location. We could not identify any significant advantage in using statistical tools such as linear regression or Bayesian networks or ad-hoc methods [15].

Uncovering Causal Relationships between Software Metrics and Bugs Bug prediction is an important challenge for software engineering research that consists in looking for possible early indicators of the presence of bugs in a software. However, despite the relevance of the issue, most experiments designed to evaluate bug prediction only investigate whether there is a linear relation between the predictor and the presence of bugs. However, it is well known that standard regression models can not filter out spurious relations. We describe an experiment to discover more robust evidences towards causality between software metrics (as predictors) and the occurrence of bugs. For this purpose, we have relied on Granger Causality Test to evaluate whether past changes in a given time series are useful to forecast changes in another series. As its name suggests, Granger Test is a better indication of causality between two variables. We present and discuss the results of experiments on four real world systems evaluated over a time frame of almost four years. Particularly, we have been able to discover in the history of metrics the causes - in the terms of the Granger Test - for 64% to 93% of the defects reported for the systems considered in our experiment [18].

BugMaps: A Tool for the Visual Exploration and Analysis of Bugs To harness the complexity of big legacy software, software engineering tools need more and more information on these systems. This information may come from analysis of the source code, study of execution traces, computing of metrics, etc. One source of information received less attention than source code: the bugs on the system. Little is known about the evolutionary behavior, lifetime, distribution, and stability of bugs. We propose to consider bugs as first class entities and a useful source of information that can answer such topics. Such analysis is inherently complex,

because bugs are intangible, invisible, and difficult to be traced. Therefore, our tool extracts information about bugs from bug tracking systems, link this information to other software artifacts, and explore interactive visualizations of bugs that we call bug maps [19].

A Catalog of Patterns for Concept Lattice Interpretation in Software Reengineering Formal Concept Analysis (FCA) provides an important approach in software reengineering for software understanding, design anomalies detection and correction. However, FCA-based approaches have two problems: (i) they produce lattices that must be interpreted by the user according to his/her understanding of the technique and different elements of the graph; and, (ii) the lattice can rapidly become so big that one is overwhelmed by the mass of information and possibilities. We make a catalog of important patterns in concept lattices, which can allow automating the task of lattice interpretation. The approach helps the reengineer to concentrate on the task of reengineering rather than understanding a complex lattice. We provide interpretation of these patterns in a generalized manner and illustrate them on various contexts constructed from program information of different open-source systems. We also present a tool that allows automated extraction of the patterns from concept lattices [16].

5.4. Reflective

Participants: Marcus Denker [Correspondant], Stéphane Ducasse.

DynamicSchema: a lightweight persistency framework for context-oriented data management While context-oriented programming technology so far has focused mostly on behavioral adaptation, context-oriented data management has received much less attention. We make a case for the problem of context-oriented data management, using a concrete example of a mobile application. We illustrate some of the issues involved and propose a lightweight persistency framework, called DynamicSchema, that resolves some of these issues. The solution consists in a flexible reification of the database schema, as a convenient dynamic data structure that can be adapted at execution time, according to sensed context changes. Implementing our mobile application using this framework enabled us to reduce the complexity of the domain modeling layer, to facilitate the production of code with low memory footprint, and to simplify the implementation of certain scenarios related to context-dependent security concerns [17].

SEQUEL Project-Team

6. New Results

6.1. Decision-making Under Uncertainty

6.1.1. Reinforcement Learning

Transfer in Reinforcement Learning: a Framework and a Survey [56]

Transfer in reinforcement learning is a novel research area that focuses on the development of methods to transfer knowledge from a set of source tasks to a target task. Whenever the tasks are *similar*, the transferred knowledge can be used by a learning algorithm to solve the target task and significantly improve its performance (e.g., by reducing the number of samples needed to achieve a nearly optimal performance). In this chapter we provide a formalization of the general transfer problem, we identify the main settings which have been investigated so far, and we review the most important approaches to transfer in reinforcement learning.

Online Regret Bounds for Undiscounted Continuous Reinforcement Learning [44]

We derive sublinear regret bounds for undiscounted reinforcement learning in continuous state space. The proposed algorithm combines state aggregation with the use of upper confidence bounds for implementing optimism in the face of uncertainty. Beside the existence of an optimal policy which satisfies the Poisson equation, the only assumptions made are Holder continuity of rewards and transition probabilities.

Semi-Supervised Apprenticeship Learning [23]

In apprenticeship learning we aim to learn a good policy by observing the behavior of an expert or a set of experts. In particular, we consider the case where the expert acts so as to maximize an unknown reward function defined as a linear combination of a set of state features. In this paper, we consider the setting where we observe many sample trajectories (i.e., sequences of states) but only one or a few of them are labeled as experts' trajectories. We investigate the conditions under which the remaining unlabeled trajectories can help in learning a policy with a good performance. In particular, we define an extension to the max-margin inverse reinforcement learning proposed by Abbeel and Ng (2004) where, at each iteration, the max-margin optimization step is replaced by a semi-supervised optimization problem which favors classifiers separating clusters of trajectories. Finally, we report empirical results on two grid-world domains showing that the semi-supervised algorithm is able to output a better policy in fewer iterations than the related algorithm that does not take the unlabeled trajectories into account.

Fast Reinforcement Learning with Large Action Sets Using Error-Correcting Output Codes for MDP Factorization [31] [48]

The use of Reinforcement Learning in real-world scenarios is strongly limited by issues of scale. Most RL learning algorithms are unable to deal with problems composed of hundreds or sometimes even dozens of possible actions, and therefore cannot be applied to many real-world problems. We consider the RL problem in the supervised classification framework where the optimal policy is obtained through a multiclass classifier, the set of classes being the set of actions of the problem. We introduce error-correcting output codes (ECOCs) in this setting and propose two new methods for reducing complexity when using rollouts-based approaches. The first method consists in using an ECOC-based classifier as the multiclass classifier, reducing the learning complexity from O(A2) to O(Alog(A)) . We then propose a novel method that profits from the ECOC's coding dictionary to split the initial MDP into O(log(A)) separate two-action MDPs. This second method reduces learning complexity even further, from O(A2) to O(log(A)) , thus rendering problems with large action sets tractable. We finish by experimentally demonstrating the advantages of our approach on a set of benchmark problems, both in speed and performance.

Analysis of Classification-based Policy Iteration Algorithms [13]

We introduce a variant of the classification-based approach to policy iteration which uses a cost-sensitive loss function weighting each classification mistake by its actual regret, i.e., the difference between the action-value of the greedy action and of the action chosen by the classifier. For this algorithm, we provide a full finite-sample analysis. Our results state a performance bound in terms of the number of policy improvement steps, the number of rollouts used in each iteration, the capacity of the considered policy space (classifier), and a capacity measure which indicates how well the policy space can approximate policies that are greedy w.r.t. any of its members. The analysis reveals a tradeoff between the estimation and approximation errors in this classification-based policy iteration setting. Furthermore it confirms the intuition that classification-based policy iteration algorithms could be favorably compared to value-based approaches when the policies can be approximated more easily than their corresponding value functions. We also study the consistency of the algorithm when there exists a sequence of policy spaces with increasing capacity.

Minimax PAC-Bounds on the Sample Complexity of Reinforcement Learning with a Generative Model [5] [24]

We consider the problem of learning the optimal action-value function in discounted-reward Markov decision processes (MDPs). We prove new PAC bounds on the sample-complexity of two well-known model-based reinforcement learning (RL) algorithms in the presence of a generative model of the MDP: value iteration and policy iteration. The first result indicates that for an MDP with N state-action pairs and the discount factor $\gamma \in [0, 1)$ only $O(N \log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ state-transition samples are required to find an ϵ-optimal estimation of the action-value function with the probability (w.p.) $1-\delta$. Further, we prove that, for small values of ϵ, an order of $O(N \log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ samples is required to find an ϵ-optimal policy w.p. $1-\delta$. We also prove a matching lower bound of $\Omega(N \log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ on the sample complexity of estimating the optimal action-value function. To the best of our knowledge, this is the first minimax result on the sample complexity of RL: The upper bound matches the lower bound interms of N , ϵ, δ and $1/(1-\gamma)$ up to a constant factor. Also, both our lower bound and upper bound improve on the state-of-the-art in terms of their dependence on $1/(1-\gamma)$.

Optimistic planning in Markov decision processes [25]

The reinforcement learning community has recently intensified its interest in online planning methods, due to their relative independence on the state space size. However, tight near-optimality guarantees are not yet available for the general case of stochastic Markov decision processes and closed-loop, state-dependent planning policies. We therefore consider an algorithm related to $AO*$ that optimistically explores a tree representation of the space of closed-loop policies, and we analyze the near-optimality of the action it returns after n tree node expansions. While this optimistic planning requires a finite number of actions and possible next states for each transition, its asymptotic performance does not depend directly on these numbers, but only on the subset of nodes that significantly impact near-optimal policies. We characterize this set by introducing a novel measure of problem complexity, called the near-optimality exponent. Specializing the exponent and performance bound for some interesting classes of MDPs illustrates the algorithm works better when there are fewer near-optimal policies and less uniform transition probabilities.

Risk Bounds in Cost-sensitive Multiclass Classification: an Application to Reinforcement Learning [61]

We propose a computationally efficient classification-based policy iteration (CBPI) algorithm. The key idea of CBPI is to view the problem of computing the next policy in policy iteration as a classification problem. We propose a new cost-sensitive surrogate loss for each iteration of CBPI. This allows us to replace the non-convex optimization problem that needs to be solved at each iteration of the existing CBPI algorithms with a convex one. We show that the new loss is classification calibrated, and thus is a sound surrogate loss, and find a calibration function (i.e., a function that represents the convergence rate of the true loss in terms of the convergence rate of the surrogate-loss) for this loss. To the best of our knowledge, this is the first calibration result (with convergence rate) in the context of multi-class classification. As a result, we are able to extend the theoretical guarantees of the existing CBPI algorithms that deal with a non-convex optimization at each iteration to our convex and efficient algorithm, and thereby, obtain the first computationally efficient and theoretically sound CBPI algorithm.

Least-Squares Methods for Policy Iteration [55]

Approximate reinforcement learning deals with the essential problem of applying reinforcement learning in large and continuous state-action spaces, by us- ing function approximators to represent the solution. This chapter reviews least-squares methods for policy iteration, an important class of algorithms for approxi- mate reinforcement learning. We discuss three techniques for solving the core, pol- icy evaluation component of policy iteration, called: least-squares temporal difference, least-squares policy evaluation, and Bellman residual minimization. We introduce these techniques starting from their general mathematical principles and detailing them down to fully specified algorithms. We pay attention to online variants of policy iteration, and provide a numerical example highlighting the behavior of representative offline and online methods. For the policy evaluation component as well as for the overall resulting approximate policy iteration, we provide guarantees on the performance obtained asymptotically, as the number of processed samples and executed iterations grows to infinity. We also provide finite-sample results, which apply when a finite number of samples and iterations is considered. Finally, we outline several extensions and improvements to the techniques and methods reviewed

On Classification-based Approximate Policy Iteration [53]

Efficient methods for tackling large reinforcement learning problems usually exploit special structure, or regularities, of the problem at hand. For example, classification-based approximate policy iteration explicitly controls the complexity of the policy space, which leads to considerable improvement in convergence speed whenever the optimal policy is easy to represent. Conventional classification-based methods, however, do not benefit from regularities of the value function, because they typically use rollout-based estimates of the action-value function. This Monte Carlo-style approach for value estimation is data-inefficient and does not generalize the estimated value function over states. We introduce a general framework for classification-based approximate policy iteration (CAPI) which exploits regularities of both the policy and the value function. Our theoretical analysis extends existing work by allowing the policy evaluation step to be performed by any reinforcement learning algorithm (including temporal-difference style methods), by handling nonparametric representations of policies, and by providing tighter convergence bounds on the estimation error of policy learning. In our experiments, instantiations of CAPI outperformed powerful purely value-based approaches.

Conservative and Greedy Approaches to Classification-based Policy Iteration [37]

The existing classification-based policy iteration (CBPI) algorithms can be divided into two categories: *direct policy iteration* (DPI) methods that directly assign the output of the classifier (the approximate greedy policy w.r.t. the current policy) to the next policy, and *conservative policy iteration* (CPI) methods in which the new policy is a mixture distribution of the current policy and the output of the classifier. The conservative policy update gives CPI a desirable feature, namely the guarantee that the policies generated by this algorithm improve at each iteration. We provide a detailed algorithmic and theoretical comparison of these two classes of CBPI algorithms. Our results reveal that in order to achieve the same level of accuracy, CPI requires more iterations, and thus, more samples than the DPI algorithm. Furthermore, CPI may converge to suboptimal policies whose performance is not better than DPI's.

A Dantzig Selector Approach to Temporal Difference Learning [36]

LSTD is a popular algorithm for value function approximation. Whenever the number of features is larger than the number of samples, it must be paired with some form of regularization. In particular, l1-regularization methods tend to perform feature selection by promoting sparsity, and thus, are well- suited for high-dimensional problems. However, since LSTD is not a simple regression algorithm, but it solves a fixed-point problem, its integration with l1-regularization is not straightforward and might come with some drawbacks (e.g., the P-matrix assumption for LASSO-TD). In this paper, we introduce a novel algorithm obtained by integrating LSTD with the Dantzig Selector. We investigate the performance of the proposed algorithm and its relationship with the existing regularized approaches, and show how it addresses some of their drawbacks.

Finite-Sample Analysis of Least-Squares Policy Iteration [14]

In this paper, we report a performance bound for the widely used least-squares policy iteration (LSPI) algorithm. We first consider the problem of policy evaluation in reinforcement learning, that is, learning the value function of a fixed policy, using the least-squares temporal-difference (LSTD) learning method, and report finite-sample analysis for this algorithm. To do so, we first derive a bound on the performance of the LSTD solution evaluated at the states generated by the Markov chain and used by the algorithm to learn an estimate of the value function. This result is general in the sense that no assumption is made on the existence of a stationary distribution for the Markov chain. We then derive generalization bounds in the case when the Markov chain possesses a stationary distribution and is β-mixing. Finally, we analyze how the error at each policy evaluation step is propagated through the iterations of a policy iteration method, and derive a performance bound for the LSPI algorithm.

Approximate Modified Policy Iteration [47]

Modified policy iteration (MPI) is a dynamic programming (DP) algorithm that contains the two celebrated policy and value iteration methods. Despite its generality, MPI has not been thoroughly studied, especially its approximation form which is used when the state and/or action spaces are large or infinite. In this paper, we propose three implementations of approximate MPI (AMPI) that are extensions of well-known approximate DP algorithms: fitted-value iteration, fitted-Q iteration, and classification-based policy iteration. We provide error propagation analyses that unify those for approximate policy and value iteration. On the last classification-based implementation, we develop a finite-sample analysis that shows that MPI's main parameter allows to control the balance between the estimation error of the classifier and the overall value function approximation.

Bayesian Reinforcement Learning [57]

This chapter surveys recent lines of work that use Bayesian techniques for reinforcement learning. In Bayesian learning, uncertainty is expressed by a prior distribution over unknown parameters and learning is achieved by computing a posterior distribution based on the data observed. Hence, Bayesian reinforcement learning distinguishes itself from other forms of reinforcement learning by explicitly maintaining a distribution over various quantities such as the parameters of the model, the value function, the policy or its gradient. This yields several benefits: a) domain knowledge can be naturally encoded in the prior distribution to speed up learning; b) the exploration/exploitation tradeoff can be naturally optimized; and c) notions of risk can be naturally taken into account to obtain robust policies.

6.1.2. Multi-arm Bandit Theory

Learning with stochastic inputs and adversarial outputs [15]

Most of the research in online learning is focused either on the problem of adversarial classification (i.e., both inputs and labels are arbitrarily chosen by an adversary) or on the traditional supervised learning problem in which samples are independent and identically distributed according to a stationary probability distribution. Nonetheless, in a number of domains the relationship between inputs and outputs may be adversarial, whereas input instances are i.i.d. from a stationary distribution (e.g., user preferences). This scenario can be formalized as a learning problem with stochastic inputs and adversarial outputs. In this paper, we introduce this novel stochastic-adversarial learning setting and we analyze its learnability. In particular, we show that in a binary classification problem over an horizon of n rounds, given a hypothesis space H with finite VC-dimension, it is possible to design an algorithm that incrementally builds a suitable finite set of hypotheses from H used as input for an exponentially weighted forecaster and achieves a cumulative regret of order $O(\sqrt{nVC(H)\log n})$ with overwhelming probability. This result shows that whenever inputs are i.i.d. it is possible to solve any binary classification problem using a finite VC-dimension hypothesis space with a sub-linear regret independently from the way labels are generated (either stochastic or adversarial). We also discuss extensions to multi-class classification, regression, learning from experts and bandit settings with stochastic side information, and application to games.

A Truthful Learning Mechanism for Multi-Slot Sponsored Search Auctions with Externalities [35]

Sponsored search auctions constitute one of the most successful applications of *microeconomic mechanisms*. In mechanism design, auctions are usually designed to incentivize advertisers to bid their truthful valuations and, at the same time, to assure both the advertisers and the auctioneer a non–negative utility. Nonetheless, in sponsored search auctions, the click–through–rates (CTRs) of the advertisers are often unknown to the auctioneer and thus standard incentive compatible mechanisms cannot be directly applied and must be paired with an effective learning algorithm for the estimation of the CTRs. This introduces the critical problem of designing a learning mechanism able to estimate the CTRs as the same time as implementing a truthful mechanism with a revenue loss as small as possible compared to an optimal mechanism designed with the true CTRs. Previous works showed that in single–slot auctions the problem can be solved using a suitable exploration–exploitation mechanism able to achieve a per–step regret of order $O(T^{-1/3})$ (where T is the number of times the auction is repeated). In this paper we extend these results to the general case of contextual multi–slot auctions with position– and ad–dependent externalities. In particular, we prove novel upper–bounds on the revenue loss w.r.t. to a VCG auction and we report numerical simulations investigating their accuracy in predicting the dependency of the regret on the number of rounds T, the number of slots K, and the number of advertisements n.

Regret Bounds for Restless Markov Bandits [43]

We consider the restless Markov bandit problem, in which the state of each arm evolves according to a Markov process independently of the learner's actions. We suggest an algorithm that after T steps achieves $\widetilde{O}(\sqrt{T})$ regret with respect to the best policy that knows the distributions of all arms. No assumptions on the Markov chains are made except that they are irreducible. In addition, we show that index-based policies are necessarily suboptimal for the considered problem.

Online allocation and homogeneous partitioning for piecewise constant mean approximation [42]

In the setting of active learning for the multi-armed bandit, where the goal of a learner is to estimate with equal precision the mean of a finite number of arms, recent results show that it is possible to derive strategies based on finite-time confidence bounds that are competitive with the best possible strategy. We here consider an extension of this problem to the case when the arms are the cells of a finite partition P of a continuous sampling space X in Rd. Our goal is now to build a piecewise constant approximation of a noisy function (where each piece is one region of P and P is fixed beforehand) in order to maintain the local quadratic error of approximation on each cell equally low. Although this extension is not trivial, we show that a simple algorithm based on upper confidence bounds can be proved to be adaptive to the function itself in a near-optimal way, when |P| is chosen to be of minimax-optimal order on the class of alpha-Holder functions.

The Optimistic Principle applied to Games, Optimization and Planning: Towards Foundations of Monte-Carlo Tree Search [17]

This work covers several aspects of the optimism in the face of uncertainty principle applied to large scale optimization problems under finite numerical budget. The initial motivation for the research reported here originated from the empirical success of the so-called Monte-Carlo Tree Search method popularized in computer-go and further extended to many other games as well as optimization and planning problems. Our objective is to contribute to the development of theoretical foundations of the field by characterizing the complexity of the underlying optimization problems and designing efficient algorithms with performance guarantees. The main idea presented here is that it is possible to decompose a complex decision making problem (such as an optimization problem in a large search space) into a sequence of elementary decisions, where each decision of the sequence is solved using a (stochastic) multi-armed bandit (simple mathematical model for decision making in stochastic environments). This so-called hierarchical bandit approach (where the reward observed by a bandit in the hierarchy is itself the return of another bandit at a deeper level) possesses the nice feature of starting the exploration by a quasi-uniform sampling of the space and then focusing progressively on the most promising area, at different scales, according to the evaluations observed so far, and eventually performing a local search around the global optima of the function. The performance of the method is assessed in terms of the optimality of the returned solution as a function of the number of function evaluations. Our main contribution to the field of function optimization is a class of hierarchical optimistic algorithms designed for general search spaces (such as metric spaces, trees, graphs, Euclidean spaces, ...) with

different algorithmic instantiations depending on whether the evaluations are noisy or noiseless and whether some measure of the "smoothness" of the function is known or unknown. The performance of the algorithms depend on the local behavior of the function around its global optima expressed in terms of the quantity of near-optimal states measured with some metric. If this local smoothness of the function is known then one can design very efficient optimization algorithms (with convergence rate independent of the space dimension), and when it is not known, we can build adaptive techniques that can, in some cases, perform almost as well as when it is known.

Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation [6]

We consider optimal sequential allocation in the context of the so-called stochastic multi-armed bandit model. We describe a generic index policy, in the sense of Gittins (1979), based on upper confidence bounds of the arm payoffs computed using the Kullback-Leibler divergence. We consider two classes of distributions for which instances of this general idea are analyzed: The kl-UCB algorithm is designed for one-parameter exponential families and the empirical KL-UCB algorithm for bounded and finitely supported distributions. Our main contribution is a unified finite-time analysis of the regret of these algorithms that asymptotically matches the lower bounds of Lai and Robbins (1985) and Burnetas and Katehakis (1996), respectively. We also investigate the behavior of these algorithms when used with general bounded rewards, showing in particular that they provide significant improvements over the state-of-the-art.

Minimax strategy for Stratified Sampling for Monte Carlo [8]

We consider the problem of stratified sampling for Monte-Carlo integration. We model this problem in a multi-armed bandit setting, where the arms represent the strata, and the goal is to estimate a weighted average of the mean values of the arms. We propose a strategy that samples the arms according to an upper bound on their standard deviations and compare its estimation quality to an ideal allocation that would know the standard deviations of the strata. We provide two pseudo-regret analyses: a distribution-dependent bound of order $O(n^{-3/2})$ that depends on a measure of the disparity of the strata, and a distribution-free bound $O(n^{-4/3})$ that does not. We also provide the first problem independent (minimax) lower bound for this problem and demonstrate that MC-UCB matches this lower bound both in terms of number of samples n and in terms of number of strata K. Finally, we link the pseudo-regret with the difference between the mean squared error on the estimated weighted average of the mean values of the arms, and the optimal oracle strategy: this provides us also with a problem dependent and a problem independent rate for this measure of performance and, as a corollary, asymptotic optimality.

Upper-Confidence-Bound Algorithms for Active Learning in Multi-Armed Bandits [7]

In this paper, we study the problem of estimating uniformly well the mean values of several distributions given a finite budget of samples. If the variance of the distributions were known, one could design an optimal sampling strategy by collecting a number of independent samples per distribution that is proportional to their variance. However, in the more realistic case where the distributions are not known in advance, one needs to design adaptive sampling strategies in order to select which distribution to sample from according to the previously observed samples. We describe two strategies based on pulling the distributions a number of times that is proportional to a high-probability upper-confidence-bound on their variance (built from previous observed samples) and report a finite-sample performance analysis on the excess estimation error compared to the optimal allocation. We show that the performance of these allocation strategies depends not only on the variances but also on the full shape of the distributions.

Bandit Algorithms boost motor-task selection for Brain Computer Interfaces [32] [10]

Brain-computer interfaces (BCI) allow users to "communicate" with a computer without using their muscles. BCI based on sensori-motor rhythms use imaginary motor tasks, such as moving the right or left hand, to send control signals. The performances of a BCI can vary greatly across users but also depend on the tasks used, making the problem of appropriate task selection an important issue. This study presents a new procedure to automatically select as fast as possible a discriminant motor task for a brain-controlled button. We develop for this purpose an adaptive algorithm, *UCB-classif*, based on the stochastic bandit theory. This shortens the training stage, thereby allowing the exploration of a greater variety of tasks. By not wasting time on

inefficient tasks, and focusing on the most promising ones, this algorithm results in a faster task selection and a more efficient use of the BCI training session. Comparing the proposed method to the standard practice in task selection, for a fixed time budget, *UCB-classif* leads to an improved classification rate, and for a fixed classification rate, to a reduction of the time spent in training by 50%.

Adaptive Stratified Sampling for Monte-Carlo integration of Differentiable functions [26]

We consider the problem of adaptive stratified sampling for Monte Carlo integration of a differentiable function given a finite number of evaluations to the function. We construct a sampling scheme that samples more often in regions where the function oscillates more, while allocating the samples such that they are well spread on the domain (this notion shares similitude with low discrepancy). We prove that the estimate returned by the algorithm is almost similarly accurate as the estimate that an optimal oracle strategy (that would know the variations of the function *everywhere*) would return, and provide a finite-sample analysis.

Risk-Aversion in Multi-Armed Bandits [46]

In stochastic multi-armed bandits the objective is to solve the exploration-exploitation dilemma and ultimately maximize the expected reward. Nonetheless, in many practical problems, maximizing the expected reward is not the most desirable objective. In this paper, we introduce a novel setting based on the principle of risk-aversion where the objective is to compete against the arm with the best risk-return trade-off. This setting proves to be intrinsically more difficult than the standard multi-arm bandit setting due in part to an exploration risk which introduces a regret associated to the variability of an algorithm. Using variance as a measure of risk, we introduce two new algorithms, we investigate their theoretical guarantees, and we report preliminary empirical results.

Bandit Theory meets Compressed Sesing for high dimensional Stochastic Linear Bandit [27]

We consider a linear stochastic bandit problem where the dimension K of the unknown parameter θ is larger than the sampling budget n. In such cases, it is in general impossible to derive sub-linear regret bounds since usual linear bandit algorithms have a regret in $O(K\sqrt{n})$. In this paper we assume that θ is S-sparse, i.e. has at most S non-zero components, and that the space of arms is the unit ball for the L_2 norm. We combine ideas from Compressed Sensing and Bandit Theory and derive an algorithm with a regret bound in $O(S\sqrt{n})$. We detail an application to the problem of optimizing a function that depends on many variables but among which only a small number of them (initially unknown) are relevant.

Thompson Sampling: an Asymptotically Optimal Finite Time Analysis [38]

The question of the optimality of Thompson Sampling for solving the stochastic multi-armed bandit problem had been open since 1933. In this paper we answer it positively for the case of Bernoulli rewards by providing the first finite-time analysis that matches the asymptotic rate given in the Lai and Robbins lower bound for the cumulative regret. The proof is accompanied by a numerical comparison with other optimal policies, experiments that have been lacking in the literature until now for the Bernoulli case.

Regret bounds for Restless Markov Bandits [43]

We consider the restless Markov bandit problem, in which the state of each arm evolves according to a Markov process independently of the learner's actions. We suggest an algorithm that after T steps achieves $O(\sqrt{T})$ regret with respect to the best policy that knows the distributions of all arms. No assumptions on the Markov chains are made except that they are irreducible. In addition, we show that index-based policies are necessarily suboptimal for the considered problem.

Minimax number of strata for online Stratified Sampling given Noisy Samples [28]

We consider the problem of online stratified sampling for Monte Carlo integration of a function given a finite budget of n noisy evaluations to the function. More precisely we focus on the problem of choosing the number of strata K as a function of the budget n. We provide asymptotic and finite-time results on how an oracle that has access to the function would choose the number of strata optimally. In addition we prove a lower bound on the learning rate for the problem of stratified Monte-Carlo. As a result, we are able to state, by improving the bound on its performance, that algorithm MC-UCB, is minimax optimal both in terms of the number of samples n and the number of strata K, up to a $\log(nK)$ factor. This enables to deduce a minimax optimal bound on the difference between the performance of the estimate output by MC-UCB, and the performance of the estimate output by the best oracle static strategy, on the class of Holder continuous functions, and up to a factor $\log(n)$.

Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence [33]

We study the problem of identifying the best arm(s) in the stochastic multi-armed bandit setting. This problem has been studied in the literature from two different perspectives: fixed budget and fixed confidence. We propose a unifying approach that leads to a meta-algorithm called unified gap-based exploration (UGapE), with a common structure and similar theoretical analysis for these two settings. We prove a performance bound for the two versions of the algorithm showing that the two problems are characterized by the same notion of complexity. We also show how the UGapE algorithm as well as its theoretical analysis can be extended to take into account the variance of the arms and to multiple bandits. Finally, we evaluate the performance of UGapE and compare it with a number of existing fixed budget and fixed confidence algorithms.

6.2. Statistical Analysis of Time Series

6.2.1. Prediction of Sequences of Structured and Unstructured Data

Reducing statistical time-series problems to binary classification [45]

We show how binary classification methods developed to work on i.i.d. data can be used for solving statistical problems that are seemingly unrelated to classification and concern highly-dependent time series. Specifically, the problems of time-series clustering, homogeneity testing and the three-sample problem are addressed. The algorithms that we construct for solving these problems are based on a new metric between time-series distributions, which can be evaluated using binary classification methods. Universal consistency of the proposed algorithms is proven under most general assumptions. The theoretical results are illustrated with experiments on synthetic and real-world data.

6.2.2. Hypothesis Testing

Testing composite hypotheses about discrete ergodic processes [21]

Given a discrete-valued sample $X_1, \cdots, X_n$ we wish to decide whether it was generated by a distribution belonging to a family H_0, or it was generated by a distribution belonging to a family H_1. In this work we assume that all distributions are stationary ergodic, and do not make any further assumptions (in particular, no independence or mixing rate assumptions). We find some necessary and some sufficient conditions, formulated in terms of the topological properties of H_0 and H_1, for the existence of a consistent test. For the case when H_1 is the complement of H_0 (to the set of all stationary ergodic processes) these necessary and sufficient conditions coincide, thereby providing a complete characterization of families of processes membership to which can be consistently tested, against their complement, based on sampling. This criterion includes as special cases several known and some new results on testing for membership to various parametric families, as well as testing identity, independence, and other hypotheses.

Uniform hypothesis testing for finite-valued stationary processes [22]

Given a discrete-valued sample $X_1, \cdots, X_n$ we wish to decide whether it was generated by a distribution belonging to a family H_0, or it was generated by a distribution belonging to a family H_1. In this work we assume that all distributions are stationary ergodic, and do not make any further assumptions (e.g. no independence or mixing rate assumptions). We would like to have a test whose probability of error (both Type I and Type II) is uniformly bounded. More precisely, we require that for each ϵ there exist a sample size n such that probability of error is upper-bounded by ϵ for samples longer than n. We find some necessary and some sufficient conditions on H_0 and H_1 under which a consistent test (with this notion of consistency) exists. These conditions are topological, with respect to the topology of distributional distance.

6.2.3. Change Point Analysis

Locating Changes in Highly Dependent Data with Unknown Number of Change Points [39]

The problem of multiple change point estimation is considered for sequences with unknown number of change points. A consistency framework is suggested that is suitable for highly dependent time-series, and an asymptotically consistent algorithm is proposed. In order for the consistency to be established the only assumption required is that the data is generated by stationary ergodic time-series distributions. No modeling, independence or parametric assumptions are made; the data are allowed to be dependent and the dependence can be of arbitrary form. The theoretical results are complemented with experimental evaluations.

6.2.4. Clustering Time Series, Online and Offline

Online Clustering of Processes [40]

The problem of online clustering is considered in the case where each data point is a sequence generated by a stationary ergodic process. Data arrive in an online fashion so that the sample received at every time-step is either a continuation of some previously received sequence or a new sequence. The dependence between the sequences can be arbitrary. No parametric or independence assumptions are made; the only assumption is that the marginal distribution of each sequence is stationary and ergodic. A novel, computationally efficient algorithm is proposed and is shown to be asymptotically consistent (under a natural notion of consistency). The performance of the proposed algorithm is evaluated on simulated data, as well as on real datasets (motion classification).

Incremental Spectral Clustering with the Normalised Laplacian [52]

Partitioning a graph into groups of vertices such that those within each group are more densely connected than vertices assigned to different groups, known as graph clustering, is often used to gain insight into the organization of large scale networks and for visualization purposes. Whereas a large number of dedicated techniques have been recently proposed for static graphs, the design of on-line graph clustering methods tailored for evolving networks is a challenging problem, and much less documented in the literature. Motivated by the broad variety of applications concerned, ranging from the study of biological networks to graphs of scientific references through to the exploration of communications networks such as the World Wide Web, it is the main purpose of this paper to introduce a novel, computationally efficient, approach to graph clustering in the evolutionary context. Namely, the method promoted in this article is an incremental eigenvalue solution for the spectral clustering method described by Ng. et al. (2001). Beyond a precise description of its practical implementation and an evaluation of its complexity, its performance is illustrated through numerical experiments, based on datasets modelling the evolution of a HIV epidemic and the purchase history graph of an e-commerce website.

6.2.5. Online Semi-Supervised Learning

Learning from a Single Labeled Face and a Stream of Unlabeled Data [41]

Face recognition from a single image per person is a challenging problem because the training sample is extremely small. We consider a variation of this problem. In our problem, we recognize only one person, and there are no labeled data for any other person. This setting naturally arises in authentication on personal computers and mobile devices, and poses additional challenges because it lacks negative examples. We formalize our problem as one-class classification, and propose and analyze an algorithm that learns a non-parametric model of the face from a single labeled image and a stream of unlabeled data. In many domains, for instance when a person interacts with a computer with a camera, unlabeled data are abundant and easy to utilize. This is the first paper that investigates how these data can help in learning better models in the single-image-per-person setting. Our method is evaluated on a dataset of 43 people and we show that these people can be recognized 90% of time at nearly zero false positives. This recall is 25+% higher than the recall of our best performing baseline. Finally, we conduct a comprehensive sensitivity analysis of our algorithm and provide a guideline for setting its parameters in practice.

6.3. Statistical Learning and Bayesian Analysis

6.3.1. Non-parametric Methods for Function Approximation

Linear Regression with Random Projections [16]

We investigate a method for regression that makes use of a randomly generated subspace G_P (of finite dimension P) of a given large (possibly infinite) dimensional function space F, for example, $L_2([0,1]^d)$. G_P is defined as the span of P random features that are linear combinations of a basis functions of F weighted by random Gaussian i.i.d. coefficients. We show practical motivation for the use of this approach, detail the link that this random projections method share with RKHS and Gaussian objects theory and prove, both in deterministic and random design, approximation error bounds when searching for the best regression function in G_P rather than in F, and derive excess risk bounds for a specific regression algorithm (least squares regression in G_P). This paper stresses the motivation to study such methods, thus the analysis developed is kept simple for explanations purpose and leaves room for future developments.

6.3.2. Nonparametric Bayesian Estimation

DPM pour l'inférence dans les modèles dynamiques non linéaires avec des bruits de mesure alpha-stable [50]

Stable random variables are often use to model impulsive noise; Recently it has be shown that communication at very high frequency suffer from such a noise. Stable noise cannot however be considered as usual noise in estimation processes because the variance does not usually exists nor an analytic expression for the probability density function. In this work we show how to manage such a problem using a bayesian nonparametric approach. We develop a Sequential Monte Carlo based algorithm to realize the estimation in a non linear dynamical system. The measurement noise is a non-stationnary stable process and it is modeled using a Dirichlet Process Mixture.

6.3.3. Random Finite Sets for Multisensor Multitarget Tracking

Multi-sensor PHD filtering with application to sensor management [2]

The aim of multi-object filtering is to address the multiple target detection and/or tracking problem. This thesis focuses on the Probability Hypothesis Density (PHD) filter, a well-known tractable approximation of the Random Finite Set (RFS) filter when the observation process is realized by a single sensor. The first part proposes the rigorous construction of the exact multi-sensor PHD filter and its simplified expression, without approximation, through a joint partitioning of the target state space and the sensors. With this new method, the exact multi-sensor PHD can be propagated in simple surveillance scenarii. The second part deals with the sensor management problem in the PHD framework. At each iteration, the Balanced Explorer and Tracker (BET) builds a prediction of the posterior multi-sensor PHD thanks to the Predicted Ideal Measurement Set (PIMS) and produces a multi-sensor control according to a few simple operational principles adapted to surveillance activities

6.4. Applications

6.4.1. Signal Processing

Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons [19]

In global positioning systems (GPS), classical localization algorithms assume, when the signal is received from the satellite in line-of-sight (LOS) environment, that the pseudorange error distribution is Gaussian. Such assumption is in some way very restrictive since a random error in the pseudorange measure with an unknown distribution form is always induced in constrained environments especially in urban canyons due to multipath/masking effects. In order to ensure high accuracy positioning, a good estimation of the observation error in these cases is required. To address this, an attractive flexible Bayesian nonparametric noise model based on Dirichlet process mixtures (DPM) is introduced. Since the considered positioning problem involves elements of non-Gaussianity and nonlinearity and besides, it should be processed on-line, the suitability of the proposed modeling scheme in a joint state/parameter estimation problem is handled by an efficient Rao-Blackwellized particle filter (RBPF). Our approach is illustrated on a data analysis task dealing with joint estimation of vehicles positions and pseudorange errors in a global navigation satellite system (GNSS)-based localization context where the GPS information may be inaccurate because of hard reception conditions.

Dislocation detection in field environments: A belief functions contribution [20]

Dislocation is defined as the change between discrete sequential locations of critical items in field environments such as large construction projects. Dislocations on large sites of materials and critical items for which discrete time position estimates are available represent critical state changes. The ability to detect dislocations automatically for tens of thousands of items can ultimately improve project performance significantly. Detecting these dislocations in a noisy information environment where low cost radio frequency identification tags are attached to each piece of material, and the material is moved sometimes only a few meters, is the main focus of this study. We propose in this paper a method developed in the frame of belief functions to detect dislocations. The belief function framework is well-suited for such a problem where both uncertainty and imprecision are inherent to the problem. We also show how to deal with the calculations. This method has been implemented in a controlled experimental setting. The results of these experiments show the ability of the proposed method to detect materials dislocation over the site reliably. Broader application of this approach to both animate and inanimate objects is possible.

Towards dictionary learning from images with non Gaussian noise [29]

We address the problem of image dictionary learning from noisy images with non Gaussian noise. This problem is difficult. As a first step, we consider the extreme sparse code given by vector quantization, i.e. each pixel is finally associated to 1 single atom. For Gaussian noise, the natural solution is K-means clustering using the sum of the squares of differences between gray levels as the dissimilarity measure between patches. For non Gaussian noises (Poisson, Gamma,...), a new measure of dissimilarity between noisy patches is necessary. We study the use of the generalized likelihood ratios (GLR) recently introduced by Deledalle et al. 2012 to compare non Gaussian noisy patches. We propose a K-medoids algorithm generalizing the usual Linde-Buzo-Gray K-means using the GLR based dissimilarity measure. We obtain a vector quantization which provides a dictionary that can be very large and redundant. We illustrate our approach by dictionaries learnt from images featuring non Gaussian noise, and present preliminary denoising results.

6.4.2. Medical Applications

Outlier detection for patient monitoring and alerting. [12]

We develop and evaluate a data-driven approach for detecting unusual (anomalous) patient-management decisions using past patient cases stored in electronic health records (EHRs). Our hypothesis is that a patient-management decision that is unusual with respect to past patient care may be due to an error and that it is worthwhile to generate an alert if such a decision is encountered. We evaluate this hypothesis using data obtained from EHRs of 4486 post-cardiac surgical patients and a subset of 222 alerts generated from the data. We base the evaluation on the opinions of a panel of experts. The results of the study support our hypothesis that the outlier-based alerting can lead to promising true alert rates. We observed true alert rates that ranged from 25% to 66% for a variety of patient-management actions, with 66% corresponding to the strongest outliers.

6.4.3. Web Mining

Managing advertising campaigns – an approximate planning approach [11]

We consider the problem of displaying commercial advertisements on web pages, in the "cost per click" model. The advertisement server has to learn the appeal of each type of visitor for the different advertisements in order to maximize the profit. Advertisements have constraints such as a certain number of clicks to draw, as well as a lifetime. This problem is thus inherently dynamic, and intimately combines combinatorial and statistical issues. To set the stage, it is also noteworthy that we deal with very rare events of interest, since the base probability of one click is in the order of 10^4. Different approaches may be thought of, ranging from computationally demanding ones (use of Markov decision processes, or stochastic programming) to very fast ones.We introduce NOSEED, an adaptive policy learning algorithm based on a combination of linear programming and multi-arm bandits. We also propose a way to evaluate the extent to which we have to handle the constraints (which is directly related to the computation cost). We investigate the performance of our system through simulations on a realistic model designed with an important commercial web actor.

ICML Exploration & Exploitation challenge: Keep it simple! [18]

Recommendation has become a key feature in the economy of a lot of companies (online shopping, search engines...). There is a lot of work going on regarding recommender systems and there is still a lot to do to improve them. Indeed nowadays in many companies most of the job is done by hand. Moreover even when a supposedly smart recommender system is designed, it is hard to evaluate it without using real audience which obviously involves economic issues. The ICML Exploration & Exploitation challenge is an attempt to make people propose efficient recommendation techniques and particularly focuses on limited computational resources. The challenge also proposes a framework to address the problem of evaluating a recommendation algorithm with real data. We took part in this challenge and achieved the best performances; this paper aims at reporting on this achievement; we also discuss the evaluation process and propose a better one for future challenges of the same kind.

6.4.4. Games

CLOP: Confident Local Optimization for Noisy Black-Box Parameter Tuning [30]

Artificial intelligence in games often leads to the problem of parameter tuning. Some heuristics may have coefficients, and they should be tuned to maximize the win rate of the program. A possible approach is to build local quadratic models of the win rate as a function of program parameters. Many local regression algorithms have already been proposed for this task, but they are usually not robust enough to deal automatically and efficiently with very noisy outputs and non-negative Hessians. The CLOP principle, which stands for Confident Local OPtimization, is a new approach to local regression that overcomes all these problems in a simple and efficient way. CLOP discards samples whose estimated value is confidently inferior to the mean of all samples. Experiments demonstrate that, when the function to be optimized is smooth, this method outperforms all other tested algorithms.

6.5. Other Results

Sequential approaches for learning datum-wise sparse representations [9]

In supervised classification, data representation is usually considered at the dataset level: one looks for the "best" representation of data assuming it to be the same for all the data in the data space. We propose a different approach where the representations used for classification are tailored to each datum in the data space. One immediate goal is to obtain sparse datum-wise representations: our approach learns to build a representation specific to each datum that contains only a small subset of the features, thus allowing classification to be fast and efficient. This representation is obtained by way of a sequential decision process that sequentially chooses which features to acquire before classifying a particular point; this process is learned through algorithms based on Reinforcement Learning. The proposed method performs well on an ensemble of medium-sized sparse classification problems. It offers an alternative to global sparsity approaches, and is a natural framework for sequential classification problems. The method extends easily to a whole family of sparsity-related problems which would otherwise require developing specific solutions. This is the case in particular for cost-sensitive and limited-budget classification, where feature acquisition is costly and is often performed sequentially. Finally, our approach can handle non-differentiable loss functions or combinatorial optimization encountered in more complex feature selection problems.

Multiple Operator-valued Kernel Learning [60]

Positive definite operator-valued kernels generalize the well-known notion of reproducing kernels, and are naturally adapted to multi-output learning situations. This paper addresses the problem of learning a finite linear combination of infinite-dimensional operator-valued kernels which are suitable for extending functional data analysis methods to nonlinear contexts. We study this problem in the case of kernel ridge regression for functional responses with an lr-norm constraint on the combination coefficients. The resulting optimization problem is more involved than those of multiple scalar-valued kernel learning since operator-valued kernels pose more technical and theoretical issues. We propose a multiple operator-valued kernel learning algorithm based on solving a system of linear operator equations by using a block coordinatedescent procedure. We experimentally validate our approach on a functional regression task in the context of finger movement prediction in brain-computer interfaces.

SHACRA Project-Team

6. New Results

6.1. Non-Rigid Augmented Reality for Hepatic Surgery

Hepatic resection and tumors removal approaches remains a major challenge. Despite the use of new minimally invasive techniques which has several advantages such as precision, decreased blood loss, quicker healing time and less pain, the lack of informations due to poor depth perception and direct contact lost leads the surgeons and the research groups to use Augmented Reality to overcome these issues. Augmented Reality is the visual overlay of computers-generated images over real world images. This technique can be used to overlay vessels, tumors and cutting planes performed on the pre-operative data (3D reconstruction from CT or MR scan) onto the laparoscopic video per-operatively. However, current techniques are limited to a rigid registration of the pre-operative liver anatomy onto the intra-operative image, and often this registration is not performed automatically. Our objective is to develop a real-time, non-rigid registration and tracking of the intra and pre-operative liver data.

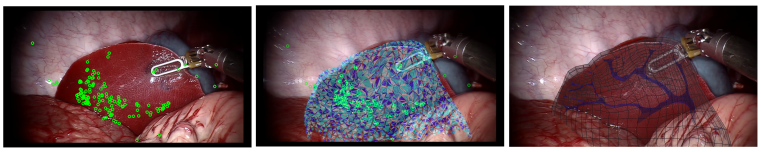


Figure 5. Non-rigid augmentation of a vascular network of a porcine liver : (left) The liver tracking. (Middle) Biomechanical model of the liver under deformation. (Right) Overlaid vascular network.

6.2. Implicit Modeling of Vascular Trees

Many clinical applications require a vessel segmentation process that is able to both extract the centerline and the surface of the blood vessels. However, noise and topology issues (such as kissing vessels) prevent existing algorithms from being able to easily retrieve such a complex system as the brain vasculature. We propose a new blood vessel tracking algorithm that 1) detect the vessel centerline; 2) provide a local radius estimate; and 3) extracts a dense set of points at the blood vessel surface. This algorithm is based on a RANSAC-based robust fitting of successive cylinders along the vessel. Our method was validated against the Multiple Hypothesis Tracking (MHT) algorithm on 10 3DRA patient data of the brain vasculature. Over 30 blood vessels of various sizes were considered for each patient. Our results demonstrated a greater ability of our algorithm to track small, tortuous and touching vessels (96% success rate), compared to MHT (65% success rate). The computed centerline precision was below 1 voxel when compared to MHT. Moreover, our results were obtained with the same set of parameters for all patients and all blood vessels, except for the seed point for each vessel, also necessary for MHT. The proposed algorithm is thereafter able to extract the full intracranial vasculature with little user interaction.

In the context of computer-based simulation, contact management requires an accurate, smooth, but still efficient surface model for the blood vessels. A new implicit model is proposed, consisting of a tree of local implicit surfaces generated by skeletons (*blobby models*). The surface is reconstructed from data points by minimizing an energy, alternating with an original blob selection and subdivision scheme. The reconstructed models are very efficient for simulation and were shown to provide a sub-voxel approximation of the vessel surface on 5 patients.

6.3. Riskmaps in DBS

As discussed in previous sections, Deep Brain Stimulation is a neurosurgical treatment that provides remarkable benefits in neurological movement and affective disorders. It consists in the implantation of a wired electrode deep into the brain. However, the accuracy of the placement is difficult due to brain shifts occuring during the procedure. Due to a potential risk of hemorrhage during the implantation, we specially investigated the brain shift induced motion of the vascular structures. We proposed a method to estimate this motion, based on a physics simulation that consider brain deformation, cerebrospinal fluid and multiple interactions, such as brain-skull contacts etc. The aim is to take it into account during the pre-operative planification step. Thus, we developped a brain-shift aware risk map. It estimate the risk for a trajectory to dissect a vessel. It could help surgeons to choose a safer trajectory for the electrode, and then avoid hemorrhages. The next steps is the use of more complex deformation models.

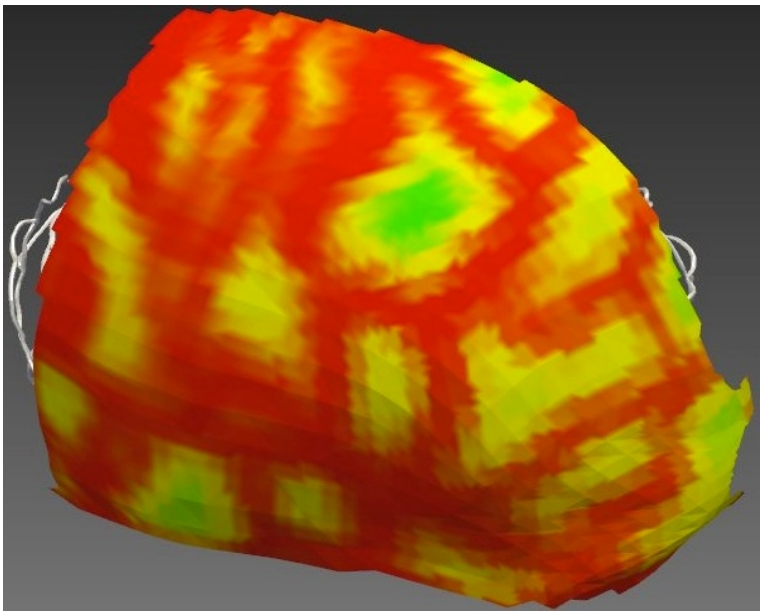


Figure 6. Brain-shift aware risk map

6.4. Electro Physiology

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. This Ph.D. aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias. The latest improvements lead on electrophysiology simulation (using GPU computing) allowed us to reach real-time performance. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.

This new result enabled us to couple the cardiac eletrophysiology with cardiac mechanical models, thus leading to an interactive framework. Our tractable simulation can therefore simulate a patient-specific electrophysiology and then compute the associated cardiac motion using an electromechanical model.

Moreover, the electrophysiology simulation has been also coupled with a navigation simulation. This is still a work in progress. The implementation of more complex models, such as bidomain models, is also in progress.

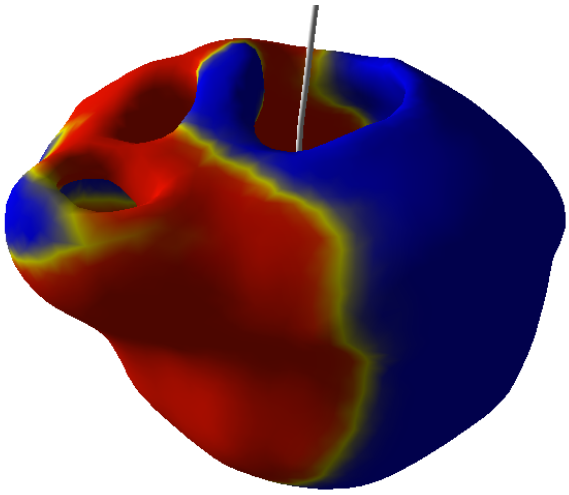


Figure 7. Cardiac electrophysiology computed on a patient-specific geometry

6.5. Shells

Many tissues in human body have thin structure and may be seen as surfaces or at least be modeled as such. Deformation modeling of surfaces is a topic with wide area of applications especially in computer graphics. However, many of the previously presented techniques are not applicable to the area of surgical simulations where a more physically based approach is desired.

To address this problem we present a new model of shell elements based on the formulation of Bézier triangles. To reduce the number of necessary degrees of freedom a kinematic link between nodes inside the element is defined. Furthermore, using implicit integration scheme allows us to achieve interactive frame rate of the simulation.

The applicability of the model has been validated on a prototype of simulator for preoperative planning of surgery of congenital heart diseases.

6.6. Interaction simulation between fluid film and deformable solids

Body fluids are a major constituent of the human body as well by their volume as by their functions. Besides the blood and the lymphatic liquid, many other liquids are present in the body and they have important functions such as lubrication or shock absorption. In this work, we are more particularly interested in the fluids being in the interface between two anatomical structures. We present a method making it possible to simulate the phenomena of interaction between a fluid film and surfaces between which it is forced. The approach that we propose is based on a fluid model and its mechanical coupling with deformable surfaces. According to the pressure of the fluid and the stiffness of the deformable solids in contact with the fluid, various behaviours are expected. Our preliminary results show that it is possible to simulate the main features of these behaviours. Furthermore, the approaches chosen for the fluid model, the deformable model and the coupling between both, are compatible with real time simulations.

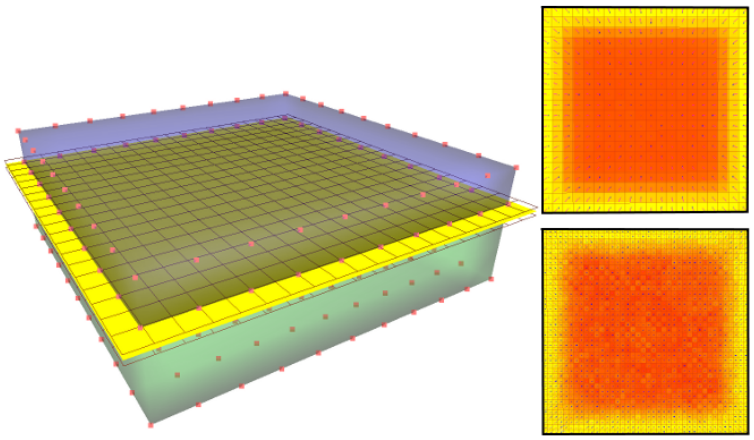


Figure 8. The fluid is between a rigid solid (green) and a deformable solid (blue). The deformable solid is constraint at the edges. Right: the height map of the fluid (yellow minimum and red maximum height).

SIMPAF Project-Team

6. New Results

6.1. Interactions of Macro- and Microscopic scales

6.1.1. Homogenization methods

We have obtained three types of results regarding the homogenization theory and its applications. The first series of results is related to nonlinear elasticity. In [44], A. Gloria has proved the convergence of a discrete model for rubber towards a nonlinear elasticity theory in collaboration with R. Alicandro and M. Cicalese. This analysis has motivated the study of a specific random point set to model the stochastic network of polymer chains, namely the random parking measure, and results have been obtained by A. Gloria and M. Penrose (University of Bath) in [42]. The numerical simulation of the model with the random parking measure has been addressed by A. Gloria, P. La Tallec and M. Vidrascu (project team REO) in [21], and the comparisons with mechanical experiments are promising, A related inverse problem is currently under investigation by M. de Buhan, A. Gloria, P. Le Tallec, and M. Vidrascu.

A second type of results concerns a quantitative theory of stochastic homogenization of discrete linear elliptic equations. A breakthrough has been obtained by A. Gloria and F. Otto (MPI Leipzig) in [63] and [24], who gave the first optimal variance estimate of the energy density of the corrector field for stochastic discrete elliptic equations. The proof makes extensive use of a spectral gap estimate and of deep elliptic regularity theory, bringing in fact the probabilistic arguments to a minimum. This analysis has enabled A. Gloria to propose efficient numerical homogenization methods, both in the discrete and continuum settings [62], [20], see the review article [33]. In [23], A. Gloria and J.-C. Mourrat has pushed the approach forward and introduced new approximation formulas for the homogenized coefficient. In [22] they have considered a more probabilistic approach and given a complete error analysis of a Monte-Carlo approximation of the homogenized coefficients in the discrete case. Work in progress concerns the generalization of the results on discrete elliptic equations to the continuum case.

The third direction of research concerns the periodic homogenization of a coupled elliptic/parabolic system arising in the modelling of nuclear waste storage. This work is in collaboration with the French agency ANDRA. A. Gloria, T. Goudon, and S. Krell have made a complete theoretical analysis of the problem, derived effective equations, and devised an efficient method to solve the effective problem numerically, based on the reduced basis approach, see [41]. This subject has been pushed forward by Z. Habibi in collaboration with ANDRA.

6.1.2. Statistical physics : molecular dynamics

In [28], the analysis of constrained molecular dynamics is proposed, with associated numerical schemes.

In [29], the pobabilistic derivation of the chemotaxis equation from the individual motion of bacteriae have been carried out. In [30], a numerical method with asymptotic variance reduction have been proposed.

6.2. Plasmas

We investigated a projective integration scheme for a kinetic equation in the limit of vanishing mean free path, in which the kinetic description approaches a diffusion phenomenon. The scheme first takes a few small steps with a simple, explicit method, such as a spatial centered flux/forward Euler time integration, and subsequently projects the results forward in time over a large time step on the diffusion time scale. We showed that, with an appropriate choice of the inner step size, the time-step restriction on the outer time step is similar to the stability condition for the diffusion equation, whereas the required number of inner steps does not depend on the mean free path. We also provided a consistency result. The presented method is asymptotic-preserving, in the sense that the method converges to a standard finite volume scheme for the diffusion equation in the limit of vanishing mean free path. This is a joint work with G. Samaey (K. U. Leuven) [27].

6.3. Finite element and finite volume methods

6.3.1. Control in fluid mechanics

Recently, open and closed active flow control were carried out in order to study the flow behavior over a backward-facing step in a transitional regime. It was done either by a global frequency destabilization at the entry of the domain, or by a local blowing or suction through the lower and upper parts of the step by the use of small jets ([58], E. Creusé, A. Giovannini (IMFT Toulouse) and I. Mortazavi (MC2 Inria EPI, Bordeaux)). The numerical computations were based on a vortex-in-cell method. Such controls were shown to be efficient in reducing the average recirculation length value, the global flow energy, as well as the global flow enstrophy. We have now in mind to apply such a strategy on cavity-stent flows, in order to study the effect of passive and/or active control on the average emptying time of the cavity, corresponding to a lot of possible industrial or health applications (combustion, blood circulation in arteries,...).
Passive as well as active control were also performed on the "Ahmed body geometry", which can be considered as a first approximation of a vehicle profile. This work was carried out in collaboration with the EPI Inria MC2 team in Bordeaux (C.H. Bruneau, I. Mortazavi and D. Depeyras), as well as with Renault car industry (P. Gillieron). We recently combined active and passive control strategies in order to reach efficient results, especially concerning the drag coefficient, for two and three dimensional simulations [51]. We recently worked on a 25° rear-window configuration of the Ahmed body, for which the 3D-effects are very important and have to be considered in the numerical simulations [9]. Moreover, the effect of the vortices dynamics on the drag coefficient of a square Ahmed body was adressed [53], as well as the impact of several Ahmed bodies on the same road [52].

In another field of applications, a work was performed with the TEMPO Laboratory of Valenciennes. The objective of this collaboration was to study the pressure wave generated by high-speed trains entering tunnels in order to improve the shape of the tunnel sections.

6.3.2. Numerical Methods for viscous flows

In the case of compressible models, as the Euler equations, a careful analysis of sharp and practical stability conditions to ensure the positivity of both density and pressure variables was performed [11]. We are also concerned with the numerical simulation of certain multi-fluids flows, which in particular arises in the modelling of powdersnow avalanches. The hybrid scheme works on unstructured meshes and can be advantageously coupled to mesh refinements strategies in order to follow fronts of high density variation [38]. In order to answer these questions, we have developed a MATLAB code (NS2DDV-M, see the softwares section), a Fortran code and a C++ code.

6.3.3. A posteriori error estimators for finite element methods

A recent work, in collaboration with S. Nicaise (LAMAV, Valenciennes), was devoted to the derivation of some so-called "reconstruction estimators" based on gradient averaging, in order to provide lower and upper bounds of the error arising from a discontinuous Galerkin approximation of a diffusion problem [59].

At the same time, some equilibrated-type estimators were developed for the Reissner-Mindlin system arising in solid mechanics applications, for conforming and locking-free approximations, in the context of the PhD. of É. Verhille.

At last, a collaboration with the "Laboratoire d'électrotechnique et d'électronique de puissance de Lille (L2EP)" began two years ago, to derive a residual-based a posteriori error estimator for the Maxwell system in its vectorial and scalar potential formulation A/Φ (PhD of Z. Tang). The objective was to obtain a mathematical rigorous error indicator, in order to couple it with the automatic mesh generator used by EDF for very practical issues.

Some residual-type a posteriori error estimators were developed in the context of magnetostatic and magnetodynamic Maxwell equations, given in their potential and harmonic formulations. Here, the task was to found a relevant decomposition of the error in order to obtain the reliability of the estimator, with the use of ad-hoc interpolations. This work was realized in collaboration with the L2EP Laboratory (Laboratoire d'Electrotechnique et d'Electronique de Puissance de Lille, Lille 1 University), and gave rise to several contributions [17], [18], [32], [31], [65], obtained in the context of the Ph-D thesis of Zuqi Tang [2]. Then, other results about reconstructed a posteriori error estimators were obtained for Discontinuous Galerkin methods, applied to convection-reaction-diffusion equations [16].

6.4. Numerical anlaysis of Schrödinger equations

6.4.1. Absorbing boundary conditions

C. Besse continues his collaboration with X. Antoine (EPI Corida) and P. Klein. They construct in [3] some classes of absorbing boundary conditions for the two-dimensional Schrödinger equation with a time and space varying exterior potential and for general convex smooth boundaries. The construction is based on asymptotics of the inhomogeneous pseudodifferential operators defining the related Dirichlet-to-Neumann operator. Furthermore, a priori estimates are developed for the truncated problems with various increasing order boundary conditions. They propose in [34] some suitable discretization schemes of these ABCs and prove some semi-discrete stability results. Furthermore, the full numerical discretization of the corresponding initial boundary value problems is considered and simulations are provided to compare the accuracy of the different ABCs.

6.4.2. Semi-classical limit of the nonlinear Schrödinger equation

C. Besse works with R. Carles and F. Méhats (EPI Ipso). They consider in [36] the semiclassical limit for the nonlinear Schrödinger equation. They introduce a phase/amplitude representation given by a system similar to the hydrodynamical formulation, whose novelty consists in including some asymptotically vanishing viscosity. They prove that the system is always locally well-posed in a class of Sobolev spaces, and globally well-posed for a fixed positive Planck constant in the one-dimensional case. They propose a second order numerical scheme which is asymptotic preserving. Before singularities appear in the limiting Euler equation, they recover the quadratic physical observables as well as the wave function with mesh size and time step independent of the Planck constant. This approach is also well suited to the linear Schrödinger equation.

6.4.3. Analysis and numerical simulation of the Schrödinger equation

The linear or nonlinear Schrödinger equation with potential is one of the basic equations of quantum mechanics and it arises in many areas of physical and technological interest, e.g. in quantum semiconductors, in electromagnetic wave propagation, and in seismic migration. The Schrödinger equation is the lowest order one-way approximation (paraxial wave equation) to the Helmholtz equation and is called Fresnel equation in optics, or standard parabolic equation in underwater acoustics. The solution of the equation is defined on an unbounded domain. If one wants to solve such a whole space evolution problem numerically, one has to restrict the computational domain by introducing artificial boundary conditions. So, the objective is to approximate the exact solution of the whole-space problem, restricted to a finite computational domain. A review article [45] was written this year to describe and compare the different current approaches of constructing and discretizing the transparent boundary conditions in one and two dimensions. However, these approaches are limited to the linear case (or nonlinear with the classical cubic nonlinearity: an article written was dedicated to this case this year [49]) and constant potentials. Therefore, in collaboration with X. Antoine (IECN Nancy and Inria Lorraine), we proposed to P. Klein to study, in her PhD thesis, the case of the Schrödinger equation with variable potentials. The study of the non-stationary one-dimensional case has already led to one publication [46] and some preliminary results in the stationary case are really promising. These cases are relevant since for example the equations appear in the Bose Einstein condensate with a quadratic potential.

This problem is obviously not limited to the Schrödinger equation and new developments are in progress on the Korteweg de Vries equation with M. Ehrhardt. This equation is more difficult to study due to its third order derivative in space.

Dispersive equations, such as the Schrödinger equation are also considered as boundary-value problems. For example, in [60], G. Dujardin studies the long time asymptotics of the solutions of linear Schrodinger equations considered as initial-boundary value problems on the half-line and on bounded intervals when the boundary data are periodic functions of time. G. Dujardin obtains theoretical results using a transformation method introduced by T. Fokas and provides several numerical experiments to support them.

6.5. Other contributions

6.5.1. Corrosion models

The Diffusion Poisson Coupled Model [47] is a model of iron based alloy in a nuclear waste repository. It describes the growth of an oxide layer in this framework. The system is made of a Poisson equation on the electrostatic potential and convection-diffusion equations on the densities f charge carriers (electrons, ferric cations and oxygen vacancies), supplemented with coupled Robin boundary conditions. The DPCM model also takes into account the growth of the oxide host lattice and its dissolution, leading to moving boundary equations. In [12], C. Chainais-Hillairet and I. Lacroix-Violet consider a simplified version of this model, where only two charge carriers are taken into account and where there is no evolution of the layer thickness. They prove the existence of a steady-state solution to this model. More recently, C. Chainais-Hillairet and I. Lacroix-Violet have also obtained an existence result for the time-dependent simplified model. This result will be soon submitted for publication.

In [4], C. Chainais-Hillairet and coworkers have studied some numerical methods for the approximation of the DPCM model. The choice of the numerical methods is justified by a stability analysis and by the study of their numerical performance. These methods have been implemented in the code CALIPSO developed at ANDRA. Numerical experiments with real-life data show the efficiency of the developed methods.

6.5.2. Transparent boundary conditions for the Helmholtz equation

C. Besse and I. Violet start a collaboration with S. Fliss (Poems), J. Coatleven and K. Ramdani (Corida) to build transparent boundary conditions for the Helmholtz equation. They propose in [6] a strategy to determine the Dirichlet-to-Neumann (DtN) operator for infinite, lossy and locally perturbed hexagonal periodic media. They obtain a factorization of this operator involving two non local operators. The first one is a DtN type operator and corresponds to a half-space problem. The second one is a Dirichlet-to-Dirichlet (DtD) type operator related to the symmetry properties of the problem. The half-space DtN operator is characterized via Floquet-Bloch transform, a family of elementary strip problems and a family of stationary Riccati equations. The DtD operator is the solution of an affine operator valued equation which can be reformulated as a non standard integral equation.

6.5.3. Analysis of subcycling techniques

Several physics situations involve phenomena which occur on very different time scales. A popular option to integrate the equations in time in this context is to use sub-cycling techniques, which allow to weaken the stability constraints. Several questions are still open for the asymptotic behavior of such methods, *e.g.* the preservation of equilibrium states. New results about the asymptotic orders if such methods have been derived on toy-model problems which allow a better understanding of these methods and their preservation of equilibrium states [40].

6.5.4. Phase transitions

We analyzed numerically a forward-backward diffusion equation with a cubic- like diffusion function, –emerging in the framework of phase transitions modelling– and its “entropy” formulation determined by considering it as the singular limit of a third-order pseudo-parabolic equation. Precisely, we proposed schemes for both the second and the third order equations, we discussed the analytical properties of their semi-discrete counter- parts and we compared the numerical results in the case of initial data of Riemann type, showing strengths and flaws of the two approaches, the main emphasis being put on the propagation of transition interfaces. This is a joint work with C. Mascia (Univ. La Sapienza) [25].

6.5.5. Modelling of the biological populations

We worked on two problems of biological populations: the understanding of the occurrence of collective behavior for large populations and the extinction probabilities in some population dynamics.

Several approaches are used in the modelling of collective behavior models for large populations of fish : we obtained results at the particle and kinetic levels for a model involving self-propulsion, friction and an attractive/repulsive potential. By introducing a new dimensionless setting, we identified five parameters that govern the possible asymptotic states for this system (clumps, spheres, dispersion, mills, rigid-body rotation, flocks) and performed a numerical analysis on the 3D particle-setting. Also, we described the kinetic system derived as the limit from the particle model as N tends to infinity; and we proposed, in 1D, a numerical scheme for the simulations, and performed a numerical analysis devoted to trying to recover asymptotically patterns similar to those emerging for the equivalent particle systems, when particles originally evolved on a circle. this is a joint work with J. Rosado (UCLA) and F. Vecil (Univ. Valencia) [43].
The extinction probabilities of a flower population may be modelled by an imhomogeneous random walk on the positive quadrant. On the one hand, introducing the generating function, that solves a PDE, we computed an explicit solution. On the other hand, we compared stochastic and deterministic resolutions of the random walk. This is a joint work with K. Raschel (Univ. Tours), V. C. Tran (Univ. Lille 1) [26].