Activity Report 2012

Section Scientific Foundations

BYMOORE Exploratory Action (section vide)

POPIX Exploratory Action

3. Scientific Foundations

3.1. Scientific Foundations

Mathematical models that characterize comple biological phenomena are complex numerical models which are defined by systems of ordinary differential equations when dealing with dynamical systems that evolve with respect to time, or by partial differential equations when there is a spatial component to the model. Also, it is sometimes useful to integrate a stochastic aspect into the dynamical systems in order to model stochastic intra-individual variability.

In order to use such methods, we are rapidly confronted with complex numerical difficulties, generally related to resolving the systems of differential equations. Furthermore, to be able to check the quality of a model, we require data. The statistical aspect of the model is thus critical in its way of taking into account different sources of variability and uncertainty, especially when data comes from several individuals and we are interested in characterizeing the inter-subject variability. Here, the tool of reference is mixed-effects models.

Mixed-effects models are statistical models with both fixed effects and random effects, i.e., mixed effects. They are useful in many real-world situations, especially in the physical, biological and social sciences. In particular, they are well-adapted to situations where repeated measurements are made on the same individual/statistical unit.

POPIX develops new methods for estimation of complex mixed-effects models. Some of the extensions to these models that POPIX is actively researching include:

- models defined by a large system of differential equations
- models defined by a system of stochastic differential equations
- mixed hidden Markov models
- mixture models and model mixtures
- time-to-event models
- models including a large number of covariates

<div align="center">

COMETE Project-Team

</div>

3. Scientific Foundations

3.1. Probability and information theory

Participants: Miguel Andrés, Nicolás Bordenabe, Konstantinos Chatzikokolakis, Ehab ElSalamouny, Sardaouna Hamadou, Catuscia Palamidessi, Marco Stronati.

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

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

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

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

3.2. The probabilistic asynchronous π-calculus

Participants: Konstantinos Chatzikokolakis, Marco Giunti, Catuscia Palamidessi, Frank Valencia, Lili Xu.

We will focus our efforts on a probabilistic variant of the asynchronous π-calculus, which is a formalism designed for mobile and distributed computation. A characteristic of our calculus is the presence of both probabilistic and nondeterministic aspects. This combination is essential to represent probabilistic algorithms and protocols, and express their properties in presence of unpredictable (nondeterministic) users and adversaries.

3.3. Expressiveness issues

Participants: Andrés Aristizábal, Catuscia Palamidessi, Luis Fernando Pino Duque, Frank Valencia.

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

3.4. Concurrent constraint programming

Participants: Andrés Aristizábal, Sophia Knight, Luis Fernando Pino Duque, Frank Valencia.

Concurrent constraint programming (ccp) is a well-established process calculus [43] for modeling systems where agents interact by adding and asking information in a global store. This information is represented as first-order logic formulae, called constraints, on the shared variables of the system (e.g., $X > 42$). The most distinctive and appealing feature of ccp is perhaps that it unifies in a single formalism the operational view of processes based upon process calculi with a declarative one based upon first-order logic. It also has an elegant denotational semantics that interprets processes as closure operators (over the set of constraints ordered by entailment). In other words, any ccp process can be seen as an idempotent, increasing, and monotonic function from stores to stores. Consequently, ccp processes can be viewed at the same time as computing agents, formulae in the underlying logic, and closure operators. This allows ccp to benefit from the large body of techniques of process calculi, logic and domain theory.

Our research in ccp develops along the following two lines:

1. The study of a bisimulation semantics for ccp. The advantage of bisimulation, over other kinds of semantics, is that it can be efficiently verified.

2. Enriching ccp with epistemic constructs, which will allow to reason about the knowledge of agents.

3.5. Model checking

Participants: Miguel Andrés, Catuscia Palamidessi.

We plan to develop model-checking techniques and tools for verifying properties of systems and protocols specified in the above formalisms.

Model checking addresses the problem of establishing whether the model (for instance, a finite-state machine) of a certain specification satisfies a certain logical formula.

We intend to concentrate our efforts on aspects that are fundamental for the verification of security protocols, and that are not properly considered in existing tools. Namely, we will focus on:

(a) the combination of probability and mobility, which is not provided by any of the current model checkers,

(b) the interplay between nondeterminism and probability, which in security present subtleties that cannot be handled with the traditional notion of scheduler,

(c) the development of a logic for expressing security (in particular privacy) properties.

Concerning the last point (the logic), we should capture both probabilistic and epistemological aspects, the latter being necessary for treating the knowledge of the adversary.

Logics of this kind have been already developed, but the investigation of the relation with the models coming from process calculi, and their utilization in model checking, is still in its infancy.

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

3. Scientific Foundations

3.1. Mesh Generation and Geometry Processing

Meshes are becoming commonplace in a number of applications ranging from engineering to multimedia through biomedecine and geology. For rendering, the quality of a mesh refers to its approximation properties. For numerical simulation, a mesh is not only required to faithfully approximate the domain of simulation, but also to satisfy size as well as shape constraints. The elaboration of algorithms for automatic mesh generation is a notoriously difficult task as it involves numerous geometric components: Complex data structures and algorithms, surface approximation, robustness as well as scalability issues. The recent trend to reconstruct domain boundaries from measurements adds even further hurdles. Armed with our experience on triangulations and algorithms, and with components from the CGAL library, we aim at devising robust algorithms for 2D, surface, 3D mesh generation as well as anisotropic meshes. Our research in mesh generation primarily focuses on the generation of simplicial meshes, i.e. triangular and tetrahedral meshes. We investigate both greedy approaches based upon Delaunay refinement and filtering, and variational approaches based upon energy functionals and associated minimizers.

The search for new methods and tools to process digital geometry is motivated by the fact that previous attempts to adapt common signal processing methods have led to limited success: Shapes are not just another signal but a new challenge to face due to distinctive properties of complex shapes such as topology, metric, lack of global parameterization, non-uniform sampling and irregular discretization. Our research in geometry processing ranges from surface reconstruction to surface remeshing through curvature estimation, principal component analysis, surface approximationand surface mesh parameterization. Another focus is on the robustness of the algorithms to defect-laden data. This focus stems from the fact that acquired geometric data obtained through measurements or designs are rarely usable directly by downstream applications. This generates bottlenecks, i.e., parts of the processing pipeline which are too labor-intensive or too brittle for practitioners. Beyond reliability and theoretical foundations, our goal is to design methods which are also robust to raw, unprocessed inputs.

3.2. Topological and Geometric Inference

Due to the fast evolution of data acquisition devices and computational power, scientists in many areas are asking for efficient algorithmic tools for analyzing, manipulating and visualizing more and more complex shapes or complex systems from approximative data. Many of the existing algorithmic solutions which come with little theoretical guarantee provide unsatisfactory and/or unpredictable results. Since these algorithms take as input discrete geometric data, it is mandatory to develop concepts that are rich enough to robustly and correctly approximate continuous shapes and their geometric properties by discrete models. Ensuring the correctness of geometric estimations and approximations on discrete data is a sensitive problem in many applications.

Data sets being often represented as point sets in high dimensional spaces, there is a considerable interest in analyzing and processing data in such spaces. Although these point sets usually live in high dimensional spaces, one often expects them to be located around unknown, possibly non linear, low dimensional shapes. These shapes are usually assumed to be smooth submanifolds or more generally compact subsets of the ambient space. It is then desirable to infer topological (dimension, Betti numbers,...) and geometric characteristics (singularities, volume, curvature,...) of these shapes from the data. The hope is that this information will help to better understand the underlying complex systems from which the data are generated. In spite of recent promising results, many problems still remain open and to be addressed, need a tight collaboration between mathematicians and computer scientists. In this context, our goal is to contribute to the development of new mathematically well founded and algorithmically efficient geometric tools for data analysis and processing of complex geometric objects. Our main targeted areas of application include machine learning, data mining, statistical analysis, and sensor networks.

3.3. Data Structures and Robust Geometric Computation

GEOMETRICA has a large expertise of algorithms and data structures for geometric problems.We are pursuing efforts to design efficient algorithms from a theoretical point of view, but we also put efforts in the effective implementation of these results.

In the past years, we made significant contributions to algorithms for computing Delaunay triangulations (which are used by meshes in the above paragraph). We are still working on the practical efficiency of existing algorithms to compute or to exploit classical Euclidean triangulations in 2 and 3 dimensions, but the current focus of our research is more aimed towards extending the triangulation efforts in several new directions of research.

One of these directions is the triangulation of non Euclidean spaces such as periodic or projective spaces, with various potential applications ranging from astronomy to granular material simulation.

Another direction is the triangulation of moving points, with potential applications to fluid dynamics where the points represent some particles of some evolving physical material, and to variational methods devised to optimize point placement for meshing a domain with a high quality elements.

Increasing the dimension of space is also a stimulating direction of research, as triangulating points in medium dimension (say 4 to 15) has potential applications and makes new challenges to trade exponential complexity of the problem in the dimension for the possibility to reach effective and practical results in reasonably small dimensions.

On the complexity analysis side, we pursue efforts to obtain complexity analysis in some practical situations involving randomized or stochastic hypotheses. On the algorithm design side, we are looking for new paradigms to exploit parallelism on modern multicore hardware architectures.

Finally, all this work is done while keeping in mind concerns related to effective implementation of our work, practical efficiency and robustness issues which have become a background task of all different works made by GEOMETRICA.

GRACE Team (section vide)

MEXICO Project-Team

3. Scientific Foundations

3.1. Concurrency

Participants: Benedikt Bollig, Thomas Chatain, Aiswarya Cyriac, Paul Gastin, Stefan Haar, Serge Haddad, Hernán Ponce de Léon, Stefan Schwoon.

Concurrency: Property of systems allowing some interacting processes to be executed in parallel.

Diagnosis: The process of deducing from a partial observation of a system aspects of the internal states or events of that system; in particular, *fault diagnosis* aims at determining whether or not some non-observable fault event has occurred.

Conformance Testing: Feeding dedicated input into an implemented system IS and deducing, from the resulting output of I, whether I respects a formal specification S.

3.1.1. Introduction

It is well known that, whatever the intended form of analysis or control, a *global* view of the system state leads to overwhelming numbers of states and transitions, thus slowing down algorithms that need to explore the state space. Worse yet, it often blurs the mechanics that are at work rather than exhibiting them. Conversely, respecting concurrency relations avoids exhaustive enumeration of interleavings. It allows us to focus on 'essential' properties of non-sequential processes, which are expressible with causal precedence relations. These precedence relations are usually called causal (partial) orders. Concurrency is the explicit absence of such a precedence between actions that do not have to wait for one another. Both causal orders and concurrency are in fact essential elements of a specification. This is especially true when the specification is constructed in a distributed and modular way. Making these ordering relations explicit requires to leave the framework of state/interleaving based semantics. Therefore, we need to develop new dedicated algorithms for tasks such as conformance testing, fault diagnosis, or control for distributed discrete systems. Existing solutions for these problems often rely on centralized sequential models which do not scale up well.

3.1.2. Diagnosis

Participants: Benedikt Bollig, Stefan Haar, César Rodríguez, Stefan Schwoon.

Fault Diagnosis for discrete event systems is a crucial task in automatic control. Our focus is on *event oriented* (as opposed to *state oriented*) model-based diagnosis, asking e.g. the following questions:
given a - potentially large - *alarm pattern* formed of observations,

- what are the possible *fault scenarios* in the system that *explain* the pattern ?
- Based on the observations, can we deduce whether or not a certain - invisible - fault has actually occurred ?

Model-based diagnosis starts from a discrete event model of the observed system - or rather, its relevant aspects, such as possible fault propagations, abstracting away other dimensions. From this model, an extraction or unfolding process, guided by the observation, produces recursively the explanation candidates.

In asynchronous partial-order based diagnosis with Petri nets [98], [99], [103], one unfolds the *labelled product* of a Petri net model $\mathcal{N}$ and an observed alarm pattern $\mathcal{A}$, also in Petri net form. We obtain an acyclic net giving partial order representation of the behaviors compatible with the alarm pattern. A recursive online procedure filters out those runs *(configurations)* that explain *exactly* $\mathcal{A}$. The Petri-net based approach generalizes to dynamically evolving topologies, in dynamical systems modeled by graph grammars, see [83].

3.1.2.1. Observability and Diagnosability

Diagnosis algorithms have to operate in contexts with low observability, i.e., in systems where many events are invisible to the supervisor. Checking *observability* and *diagnosability* for the supervised systems is therefore a crucial and non-trivial task in its own right. Analysis of the relational structure of occurrence nets allows us to check whether the system exhibits sufficient visibility to allow diagnosis. Developing efficient methods for both verification of *diagnosability checking* under concurrency, and the *diagnosis* itself for distributed, composite and asynchronous systems, is an important field for *MExICo*.

3.1.2.2. Distribution

Distributed computation of unfoldings allows one to factor the unfolding of the global system into smaller *local* unfoldings, by local supervisors associated with sub-networks and communicating among each other. In [99], [84], elements of a methodology for distributed computation of unfoldings between several supervisors, underwritten by algebraic properties of the category of Petri nets have been developed. Generalizations, in particular to Graph Grammars, are still do be done.

Computing diagnosis in a distributed way is only one aspect of a much vaster topic, that of *distributed diagnosis* (see [96], [110]). In fact, it involves a more abstract and often indirect reasoning to conclude whether or not some given invisible fault has occurred. Combination of local scenarios is in general not sufficient: the global system may have behaviors that do not reveal themselves as faulty (or, dually, non-faulty) on any local supervisor's domain (compare [82], [87]). Rather, the local diagnosers have to join all *information* that is available to them locally, and then deduce collectively further information from the combination of their views. In particular, even the *absence* of fault evidence on all peers may allow to deduce fault occurrence jointly, see [116], [117]. Automatizing such procedures for the supervision and management of distributed and locally monitored asynchronous systems is a mid-term goal of *MExICo*.

3.1.3. Verification of Concurrent Recursive Programs

Participants: Benedikt Bollig, Aiswarya Cyriac, Paul Gastin, César Rodríguez, Stefan Schwoon.

3.1.3.1. Contextual nets

Assuring the correctness of concurrent systems is notoriously difficult due to the many unforeseeable ways in which the components may interact and the resulting state-space explosion. A well-established approach to alleviate this problem is to model concurrent systems as Petri nets and analyse their unfoldings, essentially an acyclic version of the Petri net whose simpler structure permits easier analysis [97].

However, Petri nets are inadequate to model concurrent read accesses to the same resource. Such situations arise naturally in many circumstances, for instance in concurrent databases or in asynchronous circuits. The encoding tricks typically used to model these cases in Petri nets make the unfolding technique inefficient.

Contextual nets, which explicitly do model concurrent read accesses, address this problem. Their accurate representation of concurrency makes contextual unfoldings up to exponentially smaller in certain situations, which promises to yield more efficient analysis procedures. In order to realize such procedures, we shall study contextual nets and their properties, in particular the efficient construction and analysis of their unfoldings, and their applications in verification, diagnosis, and planning.

3.1.3.2. Concurrent Recursive Programs

In a DIGITEO PhD project, we will study logical specification formalisms for concurrent recursive programs. With the advent of multi-core processors, the analysis and synthesis of such programs is becoming more and more important. However, it cannot be achieved without more comprehensive formal mathematical models of concurrency and parallelization. Most existing approaches have in common that they restrict to the analysis of an over- or underapproximation of the actual program executions and do not focus on a behavioral semantics. In particular, temporal logics have not been considered. Their design and study will require the combination of prior works on logics for sequential recursive programs and concurrent finite-state programs.

3.1.4. Testing

Participants: Benedikt Bollig, Paul Gastin, Stefan Haar, Hernán Ponce de Léon.

3.1.4.1. Introduction

The gap between specification and implementation is at the heart of research on formal testing. The general *conformance testing problem* can be defined as follows: Does an implementation $\mathcal{M}'$ conform a given specification $\mathcal{M}$? Here, both $\mathcal{M}$ and $\mathcal{M}'$ are assumed to have input and output channels. The formal model $\mathcal{M}$ of the specification is entirely known and can be used for analysis. On the other hand, the implementation $\mathcal{M}'$ is unknown but interacts with the environment through observable input and output channels. So the behavior of $\mathcal{M}'$ is partially controlled by input streams, and partially observable via output streams. The Testing problem consists in computing, from the knowledge of $\mathcal{M}$, *input streams* for $\mathcal{M}'$ such that observation of the resulting output streams from $\mathcal{M}'$ allows to determine whether $\mathcal{M}'$ conforms to $\mathcal{M}$ as intended.

In this project, we focus on distributed or asynchronous versions of the conformance testing problem. There are two main difficulties. First, due to the distributed nature of the system, it may not be possible to have a unique global observer for the outcome of a test. Hence, we may need to use *local* observers which will record only *partial views* of the execution. Due to this, it is difficult or even impossible to reconstruct a coherent global execution. The second difficulty is the lack of global synchronization in distributed asynchronous systems. Up to now, models were described with I/O automata having a centralized control, hence inducing global synchronizations.

3.1.4.2. Asynchronous Testing

Since 2006 and in particular during his sabbatical stay at the University of Ottawa, Stefan Haar has been working with Guy-Vincent Jourdan and Gregor v. Bochmann of UOttawa and Claude Jard of IRISA on asynchronous testing. In the synchronous (sequential) approach, the model is described by an I/O automaton with a centralized control and transitions labeled with individual input or output actions. This approach has known limitations when inputs and outputs are distributed over remote sites, a feature that is characteristic of , e.g., web computing. To account for concurrency in the system, they have developed in [105], [88] asynchronous conformance testing for automata with transitions labeled with (finite) partial orders of I/O. Intuitively, this is a "big step" semantics where each step allows concurrency but the system is synchronized before the next big step. This is already an important improvement on the synchronous setting. The non-trivial challenge is now to cope with fully asynchronous specifications using models with decentralized control such as Petri nets.

3.1.4.3. Local Testing

Message-Sequence-Charts (MSCs) provide models of behaviors of distributed systems with communicating processes. An important problem is to test whether an implementation conforms to a specification given for instance by an HMSC. In *local testing*, one proceeds by injecting messages to the local processes and observing the responses: for each process p, a local observer records the sequence of events at p. If each local observation is consistent with some MSC defined by the specification, the implementation passes the test. If local testing on individual processes suffices to check conformance, the given specification (an HMSC language) is called locally testable. Local testability turns out to be undecidable even for regular HMSC languages [82]; the main difficulty lies in the existence of implied scenarios, i.e., global behaviors which are locally consistent with different specification scenarios. There are two approaches to attack the problem of local testing in light of this bottleneck. One is to allow joint observations of tuples of processes. This gives rise to the problem of k-testability where one allows joint observations of up to k processes [87]. We will look for structural conditions on the model or the specification ensuring k-testability. Another tactic would be to recognize that practical implementations always work with bounded buffers and impose an upper bound B on the buffer size. The set of B-bounded MSCs in the k-closure of a regular MSC language is again regular, so the B-bounded k-testability problem is decidable for all regular HMSC-definable specifications. The focus could now be on efficiently identifying the smallest k for which an HMSC specification is k-testable. Another interesting problem is to identify a minimal set of tests to validate a k-testable specification.

3.1.4.4. Goals

The first step that should be reached in the near future is the completion of asynchronous testing in the setting without any big-step synchronization. In parallel, work on the problems in local testing should progress

sufficiently to allow, in a mid-term perspective, to understand the relations and possible interconnections between local (i.e. distributed) and asynchronous (centralized) testing. This is the objective of the *TECSTES* project (2011-2014), funded by a DIGITEO *DIM/LSC* grant, and which involves Hernán Ponce de Léon and Stefan Haar of *MExICo*, and Delphine Longuet at LRI, University Paris-Sud/Orsay. We have extended several well known conformance (ioco style) relations for sequential models to models that can handle concurrency (labeled event structures). Two semantics (interleaving and partial order) were presented for every relation. With the interleaving semantics, the relations we obtained boil down to the same relations defined for labeled transition systems, since they focus on sequences of actions. The only advantage of using labeled event structures as a specification formalism for testing remains in the conciseness of the concurrent model with respect to a sequential one. As far as testing is concerned, the benefit is low since every interleaving has to be tested. By contrast, under the partial order semantics, the relations we obtain allow to distinguish explicitly implementations where concurrent actions are implemented concurrently, from those where they are interleaved, i.e. implemented sequentially. Therefore, these relations will be of interest when designing distributed systems, since the natural concurrency between actions that are performed in parallel by different processes can be taken into account. In particular, the fact of being unable to control or observe the order between actions taking place on different processes will not be considered as an impediment for testing.

In [69] and a subsequent journal submission in preparation, we develop complete testing framework for concurrent systems was developped, which included the notions of test suites and test cases. We studied what kind of systems are testable in such a framework and we have proposed sufficient conditions for obtaining a complete test suite. Finally, an algorithm to construct a test suite with such properties was proposed. These result are summarized in a paper that is being prepared for a journal submission.

The mid-to long term goal (perhaps not yet to achieve in this four-year term) is the comprehensive formalization of testing and testability in asynchronous systems with distributed architecture and test protocols.

3.2. Interaction

Participants: Benedikt Bollig, Thomas Chatain, Paul Gastin, Stefan Haar, Serge Haddad.

3.2.1. *Introduction*

Systems and services exhibit non-trivial *interaction* between specialized and heterogeneous components. This interplay is challenging for several reasons. On one hand, a coordinated interplay of several components is required, though each has only a limited, partial view of the system's configuration. We refer to this problem as *distributed synthesis* or *distributed control*. An aggravating factor is that the structure of a component might be semi-transparent, which requires a form of *grey box management*.

Interaction, one of the main characteristics of systems under consideration, often involves an environment that is not under the control of cooperating services. To achieve a common goal, the services need to agree upon a strategy that allows them to react appropriately regardless of the interactions with the environment. Clearly, the notions of opponents and strategies fall within *game theory*, which is naturally one of our main tools in exploring interaction. We will apply to our problems techniques and results developed in the domains of distributed games and of games with partial information. We will consider also new problems on games that arise from our applications.

3.2.2. *Distributed Control*

Participants: Benedikt Bollig, Thomas Chatain, Paul Gastin, Stefan Haar.

Program synthesis, as introduced by Church [95] aims at deriving directly an implementation from a specification, allowing the implementation to be correct by design. When the implementation is already at hand but choices remain to be resolved at run time then the problem becomes controller synthesis. Both program and controller synthesis have been extensively studied for sequential systems. In a distributed setting, we need to synthesize a distributed program or distributed controllers that interact locally with the system components. The main difficulty comes from the fact that the local controllers/programs have only a partial view of the entire system. This is also an old problem largely considered undecidable in most settings [114], [108], [113], [100], [102].

Actually, the main undecidability sources come from the fact that this problem was addressed in a synchronous setting using global runs viewed as sequences. In a truly distributed system where interactions are asynchronous we have recently obtained encouraging decidability results [101],[25]. This is a clear witness where concurrency may be exploited to obtain positive results. It is essential to specify expected properties directly in terms of causality revealed by partial order models of executions (MSCs or Mazurkiewicz traces). We intend to develop this line of research with the ambitious aim to obtain decidability for all natural systems and specifications. More precisely, we will identify natural hypotheses both on the architecture of our distributed system and on the specifications under which the distributed program/controller synthesis problem is decidable. This should open the way to important applications, e.g., for distributed control of embedded systems.

3.2.3. *Adaptation and Grey box management*

Participants: Benedikt Bollig, Stefan Haar, Serge Haddad.

Contrary to mainframe systems or monolithic applications of the past, we are experiencing and using an increasing number of services that are performed not by one provider but rather by the interaction and cooperation of many specialized components. As these components come from different providers, one can no longer assume all of their internal technologies to be known (as it is the case with proprietary technology). Thus, in order to compose e.g. orchestrated services over the web, to determine violations of specifications or contracts, to adapt existing services to new situations etc, one needs to analyze the interaction behavior of *boxes* that are known only through their public interfaces. For their semi-transparent-semi-opaque nature, we shall refer to them as **grey boxes**. While the concrete nature of these boxes can range from vehicles in a highway section to hotel reservation systems, the tasks of *grey box management* have universal features allowing for generalized approaches with formal methods. Two central issues emerge:

- Abstraction: From the designer point of view, there is a need for a trade-off between transparency (no abstraction) in order to integrate the box in different contexts and opacity (full abstraction) for security reasons.
- Adaptation: Since a grey box gives a partial view about the behavior of the component, even if it is not immediately useable in some context, the design of an adaptator is possible. Thus the goal is the synthesis of such an adaptator from a formal specification of the component and the environment.

3.3. Management of Quantitative Behavior

Participants: Sandie Balaguer, Benedikt Bollig, Thomas Chatain, Paul Gastin, Stefan Haar, Serge Haddad, Benjamin Monmege.

3.3.1. *Introduction*

Besides the logical functionalities of programs, the *quantitative* aspects of component behavior and interaction play an increasingly important role.

- *Real-time* properties cannot be neglected even if time is not an explicit functional issue, since transmission delays, parallelism, etc, can lead to time-outs striking, and thus change even the logical course of processes. Again, this phenomenon arises in telecommunications and web services, but also in transport systems.
- In the same contexts, *probabilities* need to be taken into account, for many diverse reasons such as unpredictable functionalities, or because the outcome of a computation may be governed by race conditions.
- Last but not least, constraints on *cost* cannot be ignored, be it in terms of money or any other limited resource, such as memory space or available CPU time.

Traditional mainframe systems were proprietary and (essentially) localized; therefore, impact of delays, unforeseen failures, etc. could be considered under the control of the system manager. It was therefore natural, in verification and control of systems, to focus on *functional* behavior entirely.

With the increase in size of computing system and the growing degree of compositionality and distribution, quantitative factors enter the stage:

- calling remote services and transmitting data over the web creates *delays*;
- remote or non-proprietary components are not "deterministic", in the sense that their behavior is uncertain.

Time and *probability* are thus parameters that management of distributed systems must be able to handle; along with both, the *cost* of operations is often subject to restrictions, or its minimization is at least desired. The mathematical treatment of these features in distributed systems is an important challenge, which *MExICo* is addressing; the following describes our activities concerning probabilistic and timed systems. Note that cost optimization is not a current activity but enters the picture in several intended activities.

3.3.2. *Probabilistic distributed Systems*

Participants: Stefan Haar, Serge Haddad.

3.3.2.1. Non-sequential probabilistic processes

Practical fault diagnosis requires to select explanations of *maximal likelihood*; this leads therefore to the question what the probability of a given partially ordered execution is. In Benveniste et al. [86], [79], we presented a model of stochastic processes, whose trajectories are partially ordered, based on local branching in Petri net unfoldings; an alternative and complementary model based on Markov fields is developed in [104], which takes a different view on the semantics and overcomes the first model's restrictions on applicability.

Both approaches abstract away from real time progress and randomize choices in *logical* time. On the other hand, the relative speed - and thus, indirectly, the real-time behavior of the system's local processes - are crucial factors determining the outcome of probabilistic choices, even if non-determinism is absent from the system.

Recently, we started a new line of research with Anne Bouillard, Sidney Rosario, and Albert Benveniste in the DistribCom team at Inria Rennes, studying the likelihood of occurrence of non-sequential runs under random durations in a stochastic Petri net setting.

Once the properties of the probability measures thus obtained are understood, it will be interesting to relate them with the two above models in logical time, and understand their differences. Another mid-term goal, in parallel, is the transfer to *diagnosis*.

3.3.2.2. Distributed Markov Decision Processes

Distributed systems featuring non-deterministic and probabilistic aspects are usually hard to analyze and, more specifically, to optimize. Furthermore, high complexity theoretical lower bounds have been established for models like partially observed Markovian decision processes and distributed partially observed Markovian decision processes. We believe that these negative results are consequences of the choice of the models rather than the intrinsic complexity of problems to be solved. Thus we plan to introduce new models in which the associated optimization problems can be solved in a more efficient way. More precisely, we start by studying connection protocols weighted by costs and we look for online and offline strategies for optimizing the mean cost to achieve the protocol. We cooperate on this subject with Eric Fabre in the DistribCom team at Inria Rennes, in the context of the DISC project.

3.3.3. *Large scale probabilistic systems*

Addressing large-scale probabilistic systems requires to face state explosion, due to both the discrete part and the probabilistic part of the model. In order to deal with such systems, different approaches have been proposed:

- Restricting the synchronization between the components as in queuing networks allows to express the steady-state distribution of the model by an analytical formula called a product-form [85].
- Some methods that tackle with the combinatory explosion for discrete-event systems can be generalized to stochastic systems using an appropriate theory. For instance symmetry based methods have been generalized to stochastic systems with the help of aggregation theory [94].

- At last simulation, which works as soon as a stochastic operational semantic is defined, has been adapted to perform statistical model checking. Roughly speaking, it consists to produce a confidence interval for the probability that a random path fulfills a formula of some temporal logic [120] .

We want to contribute to these three axes: (1) we are looking for product-forms related to systems where synchronization are more involved (like in Petri nets), (2) we want to adapt methods for discrete-event systems that require some theoretical developments in the stochastic framework and, (3) we plane to address some important limitations of statistical model checking like the expressiveness of the associated logic and the handling of rare events.

3.3.4. Real time distributed systems

Nowadays, software systems largely depend on complex timing constraints and usually consist of many interacting local components. Among them, railway crossings, traffic control units, mobile phones, computer servers, and many more safety-critical systems are subject to particular quality standards. It is therefore becoming increasingly important to look at networks of timed systems, which allow real-time systems to operate in a distributed manner.

Timed automata are a well-studied formalism to describe reactive systems that come with timing constraints. For modeling distributed real-time systems, networks of timed automata have been considered, where the local clocks of the processes usually evolve at the same rate [111] [90]. It is, however, not always adequate to assume that distributed components of a system obey a global time. Actually, there is generally no reason to assume that different timed systems in the networks refer to the same time or evolve at the same rate. Any component is rather determined by local influences such as temperature and workload.

3.3.4.1. Distributed timed systems with independently evolving clocks
Participants: Benedikt Bollig, Paul Gastin.

A first step towards formal models of distributed timed systems with independently evolving clocks was done in [80]. As the precise evolution of local clock rates is often too complex or even unknown, the authors study different semantics of a given system: The *existential semantics* exhibits all those behaviors that are possible under *some* time evolution. The *universal semantics* captures only those behaviors that are possible under *all* time evolutions. While emptiness and universality of the universal semantics are in general undecidable, the existential semantics is always regular and offers a way to check a given system against safety properties. A decidable under-approximation of the universal semantics, called *reactive semantics*, is introduced to check a system for liveness properties. It assumes the existence of a *global* controller that allows the system to react upon local time evolutions. A short term goal is to investigate a *distributed* reactive semantics where controllers are located at processes and only have local views of the system behaviors.

Several questions, however, have not yet been tackled in this previous work or remain open. In particular, we plan to exploit the power of synchronization via local clocks and to investigate the *synthesis problem*: For which (global) specifications $\mathcal{S}$ can we generate a distributed timed system with independently evolving clocks $\mathcal{A}$ (over some given system architecture) such that both the reactive and the existential semantics of $\mathcal{A}$ are precisely (the semantics of) $\mathcal{S}$? In this context, it will be favorable to have partial-order based specification languages and a partial-order semantics for distributed timed systems. The fact that clocks are not shared may allow us to apply partial-order–reduction techniques.

If, on the other hand, a system is already given and complemented with a specification, then one is usually interested in controlling the system in such a way that it meets its specification. The interaction between the actual *system* and the *environment* (i.e., the local time evolution) can now be understood as a 2-player game: the system's goal is to guarantee a behavior that conforms with the specification, while the environment aims at violating the specification. Thus, building a controller of a system actually amounts to computing winning strategies in imperfect-information games with infinitely many states where the unknown or unpredictable evolution of time reflects an imperfect information of the environment. Only few efforts have been made to tackle those kinds of games. One reason might be that, in the presence of imperfect information and infinitely many states, one is quickly confronted with undecidability of basic decision problems.

3.3.4.2. Implementation of Real-Time Concurrent Systems
Participants: Sandie Balaguer, Thomas Chatain, Stefan Haar, Serge Haddad.

This is one of the tasks of the ANR ImpRo.

The objective is to provide formal guarantees on the implementation of real-time distributed systems, despite the semantic differences between the model and the code. We consider two kinds of timed models: time Petri nets [112] and networks of timed automata [81].

Time Petri Nets allow the designer to explicit concurrent parts of the system, but without having decided yet to localize the different actions on the different components. In that sense, TPNs are more abstract than networks of timed automata, which can be seen as possible (ideal) distributed implementations. This raises the question of semantical comparison of these two models in the light of preserving the maximum of concurrency.

In order to implement our models on distributed architectures, we need a way to evaluate how much the implementation preserves the concurrency that is described in the model. For this we must be able to identify concurrency in the behavior of the models. This is done by equipping the models with a concurrent semantics (unfoldings), allowing us to consider the behaviors as partial orders.

For instance, we would like to be able to transform a time Petri net into a network of timed automata, which is closer to the implementation since the processes are well identified. But we require that this transformation preserves concurrency. Yet the first works about formal comparisons of the expressivity of these models [92], [89], [91], [93], [118] do not consider preservation of concurrency.

In contrast, we aim at formalizing and automating translations that preserve both the timed semantics and the concurrent semantics. This effort is crucial for extending concurrency-oriented methods for logical time, in particular for exploiting partial order properties. In fact, validation and management - in a broad sense - of distributed systems is not realistic *in general* without understanding and control of their real-time dependent features; the link between real-time and logical-time behaviors is thus crucial for many aspects of *MExICo*'s work.

3.3.5. *Weighted Automata and Weighted Logics*
Participants: Benedikt Bollig, Paul Gastin, Benjamin Monmege.

Time and probability are only two facets of quantitative phenomena. A generic concept of adding weights to qualitative systems is provided by the theory of weighted automata [78]. They allow one to treat probabilistic or also reward models in a unified framework. Unlike finite automata, which are based on the Boolean semiring, weighted automata build on more general structures such as the natural or real numbers (equipped with the usual addition and multiplication) or the probabilistic semiring. Hence, a weighted automaton associates with any possible behavior a weight beyond the usual Boolean classification of "acceptance" or "non-acceptance". Automata with weights have produced a well-established theory and come, e.g., with a characterization in terms of rational expressions, which generalizes the famous theorem of Kleene in the unweighted setting. Equipped with a solid theoretical basis, weighted automata finally found their way into numerous application areas such as natural language processing and speech recognition, or digital image compression.

What is still missing in the theory of weighted automata are satisfactory connections with verification-related issues such as (temporal) logic and bisimulation that could lead to a general approach to corresponding satisfiability and model-checking problems. A first step towards a more satisfactory theory of weighted systems was done in [12]. That paper, however does not give final solutions to all the aforementioned problems. It identifies directions for future research that we will be tackling.

PARSIFAL Project-Team

3. Scientific Foundations

3.1. General overview

There are two broad approaches for computational specifications. In the *computation as model* approach, computations are encoded as mathematical structures containing nodes, transitions, and state. Logic is used to *describe* these structures, that is, the computations are used as models for logical expressions. Intensional operators, such as the modals of temporal and dynamic logics or the triples of Hoare logic, are often employed to express propositions about the change in state.

The *computation as deduction* approach, in contrast, expresses computations logically, using formulas, terms, types, and proofs as computational elements. Unlike the model approach, general logical apparatus such as cut-elimination or automated deduction becomes directly applicable as tools for defining, analyzing, and animating computations. Indeed, we can identify two main aspects of logical specifications that have been very fruitful:

- *Proof normalization*, which treats the state of a computation as a proof term and computation as normalization of the proof terms. General reduction principles such as β-reduction or cut-elimination are merely particular forms of proof normalization. Functional programming is based on normalization [44], and normalization in different logics can justify the design of new and different functional programming languages [30].

- *Proof search*, which views the state of a computation as a a structured collection of formulas, known as a *sequent*, and proof search in a suitable sequent calculus as encoding the dynamics of the computation. Logic programming is based on proof search [49], and different proof search strategies can be used to justify the design of new and different logic programming languages [48].

While the distinction between these two aspects is somewhat informal, it helps to identify and classify different concerns that arise in computational semantics. For instance, confluence and termination of reductions are crucial considerations for normalization, while unification and strategies are important for search. A key challenge of computational logic is to find means of uniting or reorganizing these apparently disjoint concerns.

An important organizational principle is structural proof theory, that is, the study of proofs as syntactic, algebraic and combinatorial objects. Formal proofs often have equivalences in their syntactic representations, leading to an important research question about *canonicity* in proofs – when are two proofs "essentially the same?" The syntactic equivalences can be used to derive normal forms for proofs that illuminate not only the proofs of a given formula, but also its entire proof search space. The celebrated *focusing* theorem of Andreoli [32] identifies one such normal form for derivations in the sequent calculus that has many important consequences both for search and for computation. The combinatorial structure of proofs can be further explored with the use of *deep inference*; in particular, deep inference allows access to simple and manifestly correct cut-elimination procedures with precise complexity bounds.

Type theory is another important organizational principle, but most popular type systems are generally designed for either search or for normalization. To give some examples, the Coq system [56] that implements the Calculus of Inductive Constructions (CIC) is designed to facilitate the expression of computational features of proofs directly as executable functional programs, but general proof search techniques for Coq are rather primitive. In contrast, the Twelf system [53] that is based on the LF type theory (a subsystem of the CIC), is based on relational specifications in canonical form (*i.e.*, without redexes) for which there are sophisticated automated reasoning systems such as meta-theoretic analysis tools, logic programming engines, and inductive theorem provers. In recent years, there has been a push towards combining search and normalization in the same type-theoretic framework. The Beluga system [54], for example, is an extension of the LF type theory with a purely computational meta-framework where operations on inductively defined LF objects can be expressed as functional programs.

The Parsifal team investigates both the search and the normalization aspects of computational specifications using the concepts, results, and insights from proof theory and type theory.

3.2. Design of two level-logic systems

The team has spent a number of years in designing a strong new logic that can be used to reason (inductively and co-inductively) on syntactic expressions containing bindings. This work has been published is a series of papers by McDowell and Miller [46] [45], Tiu and Miller [51] [57], and Gacek, Miller, and Nadathur [2] [38]. Besides presenting formal properties of these logic, these papers also documented a number of examples where this logic demonstrated superior approaches to reasoning about a number of complex formal systems, ranging from programming languages to the λ-calculus and π-calculus.

The team has also been working on three different prototype theorem proving system that are all related to this stronger logic. These systems are the following.

- Abella, which is an interactive theorem prover for the full logic.
- Bedwyr, which is a model checker for the "finite" part of the logic.
- Tac, which is a sophisticate tactic for automatically completing simple proofs involving induction and unfolding.

We are now in the process of attempting to make all of these system communicate properly. Given that these systems have been authored by different team members at different times and for different reasons, they do not formally share the same notions of syntax and proof. We are now working to revisit all of these systems and revise them so that they all work on the *same* logic and so that they can share their proofs with each other.

Currently, Chaudhuri, Miller, and Accattoli are working with our technical staff member, Heath, to redesign and restructure these systems so that they can cooperate in building proofs.

3.3. Making the case for proof certificates

The team has been considering how it might be possible to define a universal format for proofs so that any existing theorem provers can have its proofs trusted by any other prover. This is a rather ambitious project and involves a great deal of work at the infrastructure level of computational logic. As a result, we have put significant energies into considering the high-level objectives and consequences of deploying such proof certificates.

Our current thinking on this point is roughly the following. Proofs, both formal and informal, are documents that are intended to circulate within societies of humans and machines distributed across time and space in order to provide trust. Such trust might lead a mathematician to accept a certain statement as true or it might help convince a consumer that a certain software system is secure. Using this general definition of proof, we have re-examined a range of perspectives about proofs and their roles within mathematics and computer science that often appears contradictory.

Given this view of proofs as both document and object, that need to be communicated and checked, we have attempted to define a particular approach to a *broad spectrum proof certificate* format that is intended as a universal language for communicating formal proofs among computational logic systems. We identify four desiderata for such proof certificates: they must be

1. checkable by simple proof checkers,
2. flexible enough that existing provers can conveniently produce such certificates from their internal evidence of proof,
3. directly related to proof formalisms used within the structural proof theory literature, and
4. permit certificates to elide some proof information with the expectation that a proof checker can reconstruct the missing information using bounded and structured proof search.

We consider various consequences of these desiderata, including how they can mix computation and deduction and what they mean for the establishment of marketplaces and libraries of proofs. More specifics can be found in Miller's papers [8] and [47].

3.4. Combining Classical and Intuitionistic Proof Systems

In order to develop an approach to proof certificates that is as comprehensive as possible, one needs to handle theorems and proofs in both classical logic and intuitionistic logic. Yet, building two separate libraries, one for each logic, can be inconvenient and error-prone. An ideal approach would be to design a single proof system in which both classical and intuitionistic proofs can exist together. Such a proof system should allow cut-elimination to take place and should have a sensible semantic framework.

Liang and Miller have recently been working on exactly that problem. In their paper [7], they showed how to describe a general setting for specifying proofs in intuitionistic and classical logic and to achieve one framework for describing initial-elimination and cut-elimination for such these two logics. That framework allowed for some mixing of classical and intuitionistic features in one logic. A more ambitious merging of these logics was provided in their work on "polarized intuitionistic logic" in which classical and intuitionistic connectives can be used within the same formulas [14].

3.5. Deep inference

Deep inference [40], [42] is a novel methodology for presenting deductive systems. Unlike traditional formalisms like the sequent calculus, it allows rewriting of formulas deep inside arbitrary contexts. The new freedom for designing inference rules creates a richer proof theory. For example, for systems using deep inference, we have a greater variety of normal forms for proofs than in sequent calculus or natural deduction systems. Another advantage of deep inference systems is the close relationship to categorical proof theory. Due to the deep inference design one can directly read off the morphism from the derivations. There is no need for a counter-intuitive translation.

The following research problems are investigated by members of the Parsifal team:

- Find deep inference system for richer logics. This is necessary for making the proof theoretic results of deep inference accessible to applications as they are described in the previous sections of this report.
- Investigate the possibility of focusing proofs in deep inference. As described before, focusing is a way to reduce the non-determinism in proof search. However, it is well investigated only for the sequent calculus. In order to apply deep inference in proof search, we need to develop a theory of focusing for deep inference.

3.6. Proof nets and atomic flows

Proof nets and atomic flows are abstract (graph-like) presentations of proofs such that all "trivial rule permutations" are quotiented away. Ideally the notion of proof net should be independent from any syntactic formalism, but most notions of proof nets proposed in the past were formulated in terms of their relation to the sequent calculus. Consequently we could observe features like "boxes" and explicit "contraction links". The latter appeared not only in Girard's proof nets [39] for linear logic but also in Robinson's proof nets [55] for classical logic. In this kind of proof nets every link in the net corresponds to a rule application in the sequent calculus.

Only recently, due to the rise of deep inference, new kinds of proof nets have been introduced that take the formula trees of the conclusions and add additional "flow-graph" information (see e.g., [4], [3] and [41]. On one side, this gives new insights in the essence of proofs and their normalization. But on the other side, all the known correctness criteria are no longer available.

This directly leads to the following research questions investigated by members of the parsifal team:

- Finding (for classical logic) a notion of proof nets that is deductive, i.e., can effectively be used for doing proof search. An important property of deductive proof nets must be that the correctness can be checked in linear time. For the classical logic proof nets by Lamarche and Straßburger [4] this takes exponential time (in the size of the net).

- Studying the normalization of proofs in classical logic using atomic flows. Although there is no correctness criterion they allow to simplify the normalization procedure for proofs in deep inference, and additionally allow to get new insights in the complexity of the normalization.

SECSI Project-Team

3. Scientific Foundations

3.1. Foundations

Computer security has become more and more pressing as a concern since the mid 1990s. There are several reasons to this: cryptography is no longer a *chasse réservée* of the military, and has become ubiquitous; and computer networks (e.g., the Internet) have grown considerably and have generated numerous opportunities for attacks and misbehaviors, notably.

The aim of the SECSI project is to *develop logic-based verification techniques for security properties of computer systems and networks*. Let us explain what this means, and what this does not mean.

First, the scope of the research at SECSI started as a rather broad subset of computer security, although the core of SECSI's activities has always been on verifying cryptographic protocols.

We took this for granted in 2006, and decided to concentrate on the latter. This already includes a vast number of concerns.

First, there is a plethora of distinct *security properties* one may wish to verify. Beyond the standard properties of secrecy (weak or strong forms), or authentication, one considers anonymity, fairness in contract-signing, and the subtle security properties involved in electronic voting such as accountability, receipt-freeness, resistance to coercion, or user verifiability. Some of these properties are trace properties, some are not, and are therefore more complex to state and verify.

Second, there are many available *models*. SECSI started with the rather simple symbolic models of security known today as Dolev-Yao models. One must then look at process algebra models (spi-calculus, applied pi-calculus), which allow for a symbolic treatment of more complex properties, especially those that are not trace properties. And one must also look at the computational models favored by cryptographers, e.g., the game-based approaches and the universal composability/simulatability approaches. They are more realistic in terms of security, but less directly amenable to automated verification. One of the features of computational models that makes them more complex is the need for computing, and bounding probabilities of certain events. This led us into contributing to the field of verification of probabilistic systems. One must also look at the relations between these models.

Third, there are many important *applications*. While SECSI started looking at the rather simple and now mundane confidentiality and authentication protocols, two important application domains have emerged: the verification of electronic voting protocols, and the verification of cryptographic APIs.

Apart from cryptographic protocols, the initial vision of the SECSI project was that computer security, being a global concern, should be taken as a whole, as far as possible. This is why one of the initial objectives of SECSI included topic in intrusion detection, again seen from the logical point of view.

One should remember the following. First, one of the key phrases in the SECSI motto is "logic-based". It is a founding theme of SECSI that logic matters in security, and opportunities are to be grabbed. Another key phrase is "verification techniques". The expertise of SECSI is not in designing protocols or security architectures. Verifying protocols, formally, is an arduous task already, and has proved to be an extremely rich area.

3.2. Objectives

SECSI has five objectives:

- Objective 1: symbolic verification of cryptographic protocols. Tree-automata based methods, automated deduction, and approximate/exact cryptographic protocol verification in the Dolev-Yao model. Enriching the Dolev-Yao model with algebraic theories, and associated decision problems.

- Objective 2: verification of cryptographic protocols in computational models. Computational soundness of formal models (Dolev-Yao, applied pi-calculus).
- Objective 3: security of group protocols, fair exchange, voting and other protocols. Other security properties, other security models. Security properties based on notions of indistinguishability.
- Objective 4: probabilistic transition systems. Security in the presence of probabilistic and demonic non-deterministic choices.
- Objective 5: intrusion detection, network and host protection in the large.

TOCCATA Team

3. Scientific Foundations

3.1. Introduction

In the former *ProVal* project, we have been working on the design of methods and tools for deductive verification of programs. One of our originalities is our ability to conduct proofs by using automatic provers and proof assistants at the same time, depending on the difficulty of the program, and specifically the difficulty of each particular verification condition. We thus believe that we are in a good position to propose a bridge between the two families of approaches of deductive verification presented above. This is a new goal of the team: we want to provide methods and tools for deductive program verification that can offer both a high amount of proof automation and a high guarantee of validity. Toward this objective, a new axis of research that we propose is to develop certified tools: analysis tools that are themselves formally proved correct.

As mentioned above, some of the members of the team have an internationally recognized expertise on deductive program verification involving floating-point computation [5], including both interactive proving and automated solving [9]. Indeed we noticed that the verification of numerical programs is a representative case that can benefit a lot from combining automatic and interactive theorem proving [62][4]. This is why certification of numerical programs is another axis of this proposition.

The *Toccata* project emphasizes two new axes of research: certified tools and verification of numerical programs. Additionally we want to continue the fundamental studies we conducted in the past concerning deductive program verification in general. This is why our detailed scientific programme is structured into three themes:

1. Certified Programs,
2. Certified Tools,
3. Certified Numerical Programs.

The reader should be aware that the word "Certified" in this scientific programme means "verified by a formal specification and a formal proof that the program meets this specification". This differs from the standard meaning of "Certified" in an industrial context which is that it conforms to a rigorous process and/or a norm.

3.2. Certified Programs

This theme of research builds upon our expertise on the development of methods and tools for proving programs, from source codes annotated with specifications to proofs. In the past years, we tackled programs written in mainstream programming languages, with the system *Why3* and the front-ends *Krakatoa* for Java source code, and Frama-C/Jessie for C code. However, Java and C programming languages were designed a long time ago, and certainly not with the objective of formal verification in mind. This raises a lot of difficulties when designing specification languages on top of them, and verification condition generators to analyze them. On the other hand, we designed and/or used the *Coq* and *Why3* languages and tools for performing deductive verification, but those were not designed as programming languages that can be compiled into executable programs.

Thus, a new axis of research we propose is the design of an environment that is aimed to both programming and proving, hence that will allow to develop certified programs. To achieve this goal, there are two major axes of theoretical research that needs to be conducted, concerning on the one hand methods required to support genericity and reusability of certified components, and on the other hand the automation of the proof of the verification conditions that will be generated.

3.2.1. Genericity and Reusability of Certified Components

A central ingredient for the success of deductive approaches in program verification is the ability to reuse components that are already proved. This is the only way to scale the deductive approach up to programs of larger size. As for programming languages, a key aspect that allow reusability is *genericity*. In programming languages, genericity typically means parametricity with respect to data types, e.g. *polymorphic types* in functional languages like ML, or *generic classes* in object-oriented languages. Such genericity features are essential for the design of standard libraries of data structures such as search trees, hash tables, etc. or libraries of standard algorithms such as for searching, sorting.

In the context of deductive program verification, designing reusable libraries needs also the design of *generic specifications* which typically involve parametricity not only with respect to data types but also with respect to other program components. For example, a generic component for sorting an array needs to be parametrized by the type of data in the array but also by the comparison function that will be used. This comparison function is thus another program component that is a parameter of the sorting component. For this parametric component, one needs to specify some requirements, at the logical level (such as being a total ordering relation), but also at the program execution level (like being *side-effect free*, i.e. comparing of data should not modify the data). Typically such a specification may require *higher-order* logic.

Another central feature that is needed to design libraries of data structures is the notion of data invariants. For example, for a component providing generic search trees of reasonable efficiency, one would require the trees to remain well-balanced, over all the life time of a program.

This is why the design of reusable certified components requires advanced features, such as *higher-order specifications and programs*, *effect polymorphism* and *specification of data invariants*. Combining such features is considered as an important challenge in the current state of the art (see e.g. [90]). The well-known proposals for solving it include *Separation logic* [107], *implicit dynamic frames* [105], and *considerate reasoning* [106]. Part of our recent research activities were aimed at solving this challenge: first at the level of specifications, e.g. we proposed generic specification constructs upon Java [108] or a system of theory cloning in our system *Why3* [1]; second at the level of programs, which mainly aims at controlling side-effects to avoid unexpected breaking of data invariants, thanks to advanced type checking : approaches based on *memory regions*, *linearity* and *capability-based* type systems [69], [88], [50].

A concrete challenge that should be solved in the future is: what additional constructions should we provide in a specification language like ACSL for C, in order to support modular development of reusable software components? In particular, what would be an adequate notion of module, that would provide a good notion of abstraction, both at the level of program components and at the level of specification components.

3.2.2. Automated Deduction for Program Verification

Verifying that a program meets formal specifications typically amounts to generate *verification conditions* e.g. using a weakest precondition calculus. These verification conditions are purely logical formulas—typically in first-order logic and involving arithmetic in integers or real numbers—that should be checked to be true. This can be done using either automatic provers or interactive proof assistants. Automatic provers do not need user interaction, but may run forever or give no conclusive answer.

There are several important issues to tackle. Of course, the main general objective is to improve automation as much as possible. We want to continue our efforts around our own automatic prover *Alt-Ergo* towards more expressivity, efficiency, and usability, in the context of program verification. More expressivity means that the prover should better support the various theories that we use for modeling. Toward this direction, we aim at designing specialized proof search strategies in *Alt-Ergo*, directed by rewriting rules, in the spirit of what we did for the theory of associativity and commutativity [6].

A key challenge is also in the better handling of quantifiers. SMT solvers, including *Alt-Ergo*, deal with quantifiers with a somewhat ad-hoc mechanism of heuristic instantiation of quantified hypotheses using the so-called *triggers* that can be given by hand [42], [32]. This is completely different from resolution-based provers of the TPTP category (E-prover, Vampire, etc.) which use unification to apply quantified premises.

A challenge is thus to find the best way to combine these two different approaches of quantifiers. Another challenge is to add some support for higher-order functions and predicates in this SMT context, since as said above, reusable certified components will require higher-order specifications. There are a few solutions that were proposed yet, that amount to encode higher-order goals in first-order ones [88].

Generally speaking, there are several theories, interesting for program verification, that we would like to add as built-in decision procedures in a SMT context. First, although there already exist decision procedures for variants of bit-vectors, these are not complete enough to support what is needed to reason on programs that manipulate data at the bit-level, in particular if conversions from bit-vectors to integers or floating-point numbers are involved [12]. Regarding floating-point numbers, an important challenge is to integrate in an SMT context a decision procedure like the one implemented in our tool *Gappa*.

Another goal is to improve the feedback given by automatic provers: failed proof attempts should be turned into potential counterexamples, so as to help debugging programs or specifications. A pragmatic goal would be to allow cooperation with other verification techniques. For instance, testing could be performed on unproved goals. Regarding this cooperation objective, an important goal is a deeper integration of automated procedures in interactive proofs, like it already exists in Isabelle [68]. We now have a prototype for a *Why3* tactic in *Coq* that we plan to improve.

3.2.3. *An environment for both programming and proving*

As said before, a new axis of research we would like to follow is to design a language and an environment for both programming and proving. We believe that this will be a fruitful approach for designing highly trustable software. This is a similar goal as projects Plaid, Trellys, ATS, or Guru, mentioned above.

The basis of this research direction is the *Why3* system, which is in fact a reimplementation from scratch of the former *Why* tool, that we started in January 2011. This new system supports our research at various levels. It is already used as an intermediate language for deductive verification.

The next step for us is to develop its use as a true programming language. Our objective is to propose a language where programs could be both executed (e.g. thanks to a compiler to, say, *OCaml*) and proved correct. The language would basically be purely applicative (i.e. without side-effects, e.g. close to ML) but incorporating specifications in its core. There are, however, some programs (e.g. some clever algorithms) where a bit of imperative programming is desirable. Thus, we want to allow some form of imperative features, but in a very controlled way: it should provide a strict form of imperative programming that is clearly more amenable to proof, in particular dealing with data invariants on complex data structures.

As already said before, reusability is a key issue. Our language should propose some form of modules with interfaces abstracting away implementation details. Our plan is to reuse the known ideas of *data refinement* [99] that was the foundation of the success of the B method. But our language will be less constrained than what is usually the case in such a context, in particular regarding the possibility of sharing data, and the constraints on composition of modules, there will be a need for advanced type systems like those based on regions and permissions.

The development of such a language will be the basis of the new theme regarding the development of certified tools, that is detailed in Section 3.3 below.

3.2.4. *Extra Exploratory Axes of Research*

In this theme concerning certified programs, there are a few extra exploratory topics that we plan to explore.

Concurrent Programming So far, we only investigated the verification of sequential programs. However, given the spreading of multi-core architectures nowadays, it becomes important to be able to verify concurrent programs. This is known to be a major challenge. We plan to investigate in this direction, but in a very careful way. We believe that the verification of concurrent programs should be done only under restrictive conditions on the possible interleaving of processes. In particular, the access and modification of shared data should be constrained by the programming paradigm, to allow reasonable formal specifications. In this matter, the issues are close to the ones about sharing data between components in sequential programs, and there are already some successful approaches like separation logic, dynamic frames, regions, and permissions.

Resource Analysis The deductive verification approaches are not necessarily limited to functional behavior of programs. For example, a formal termination proof typically provides a bound on the time complexity of the execution. Thus, it is potentially possible to verify resources consumption in this way, e.g. we could prove WCET (Worst Case Execution Times) of programs. Nowadays, WCET analysis is typically performed by abstract interpretation, and is applied on programs with particular shape (e.g. no unbounded iteration, no recursion). Applying deductive verification techniques in this context could allow to establish good bounds on WCET for more general cases of programs.

Other Programming Paradigms We are interested in the application of deductive methods in other cases than imperative programming à la C, Java or Ada. Indeed, in the recent years, we applied proof techniques to randomized programs [45], to cryptographic programs [49]. We plan to use proof techniques on applications related to databases. We also have plans to support low-level programs such as assembly code [81], [102] and other unstructured programming paradigm.

We are also investigating more and more applications of SMT solving, e.g. in model-checking approach (for example in Cubicle ¹ [24]) or abstract interpretation techniques (new project Cafein that will start in 2013) and also for discharging proof obligations coming from other systems like Atelier B [29] (new project BWare, Section 8.2.1).

3.3. Certified Tools

The goal of this theme is to guarantee the soundness of the tools we develop. Indeed it goes beyond that: our goal is to promote our future *Why3* environment so that *others* could develop certified tools. Tools like automated provers or program analysers are good candidate case studies because they are mainly performing symbolic computations, and as such they are usually programmed in a mostly purely functional style.

We conducted several experiments of development of certified software in the past. First, we have a strong expertise in the development of *libraries* in *Coq*: the Coccinelle library [74] formalizing term rewriting systems, the Alea library [45] for the formalization of randomized algorithms, several libraries formalizing floating-point numbers (Floats [58], Gappalib [97], and now Flocq [5] which unifies the formers). Second we recently conducted the development of a certified decision procedure [93] that corresponds to a core part of *Alt-Ergo*, and a certified verification condition generator for a language [28] similar to *Why*. On-going work aims at building, still in *Coq*, a certified VC generator for C annotated in ACSL [55], based on the operational semantics formalized in the CompCert certified compiler project [92].

To go further, we have several directions of research in mind.

3.3.1. Formalization of Binders

Using the *Why3* programming language instead of *Coq* allows more freedom. For example, it should allow one to use a bit of side-effects when the underlying algorithm justify it (e.g. hash-consing, destructive unification). On the other hand, we will lose some *Coq* features like dependent types that are usually useful when formalizing languages. Among the issues that should be studied, we believe that the question of the formalization of binders is both central and challenging (as exemplified by the POPLmark international challenge [47]).

The support of binders in *Why3* should not be built-in, but should be under the form of a reusable *Why3* library, that should already contain a lot of proved lemmas regarding substitution, alpha-equivalence and such. Of course we plan to build upon the former experiments done for the POPLmark challenge. Although, it is not clear yet that the support of binders only via a library will be satisfactory. We may consider addition of built-in constructs if this shows useful. This could be a form of (restricted) dependent types as in *Coq*, or subset types as in PVS.

3.3.2. Theory Realizations, Certification of Transformations

As an environment for both programming and proving, *Why3* should come with a standard library that includes both certified libraries of programs, but also libraries of specifications (e.g. theories of sets, maps, etc.).

---

¹ http://cubicle.lri.fr/

The certification of those *Why3* libraries of specifications should be addressed too. *Why3* libraries for specifying models of programs are commonly expressed using first-order axiomatizations, which have the advantage of being understood by many different provers. However, such style of formalization does not offer strong guarantees of consistency. More generally, the fact that we are calling different kind of provers to discharge our verification conditions raises several challenges for certification: we typically apply various transformations to go from the *Why3* language to those of the provers, and these transformations should be certified too.

A first attempt in considering such an issue was done in [29]. It was proposed to certify the consistency of a library of specification using a so-called *realization*, which amounts to "implementing" the library in a proof assistant like *Coq*. This will be an important topic of the new ANR project BWare.

3.3.3. *Certified Theorem Proving*

The goal is to develop *certified* provers, in the sense that they are proved to give a correct answer. This is an important challenge since there have been a significant amount of soundness bugs discovered in the past, in many tools of this kind.

The former work on the certified core of *Alt-Ergo* [93] should be continued to support more features: more theories (full integer arithmetic, real arithmetic, arrays, etc.), quantifiers. Development of a certified prover that supports quantifiers should build upon the previous topic about binders.

In a similar way, the *Gappa* prover which is specialized to solving constraints on real numbers and floating-point numbers should be certified too. Currently, *Gappa* can be asked to produce a *Coq* proof of its given goal, so as to check *a posteriori* its soundness. Indeed, the idea of producing a trace is not contradictory with certifying the tool. For very complex decision procedures, the goal of developing a certified proof search might be too ambitious, and the production of an internal trace is a general technique that might be used as a workaround: it suffices to instrument the proof search and to develop a certified trace checker to be used by the tool before it gives an answer. We used this approach in the past for certified proofs of termination of rewriting systems [75]. This is also a technique that is used internally in CompCert for some passes of compilation [92].

3.3.4. *Certified VC generation*

The other kind of tools that we would like to certify are the VC generators. This will be a continuation of the on-going work on developing in *Coq* a certified VC generator for C code annotated in ACSL. We would like to develop such a generator in *Why3* instead of *Coq*. As before, this will build upon a formalization of binders.

There are various kinds of VC generators that are interesting. A generator for a simple language in the style of those of *Why3* is a first step. Other interesting cases are: a generator implementing the so-called *fast weakest preconditions* [91], and a generator for unstructured programs like assembly, that would operate on an arbitrary control-flow graph.

On a longer term, it would be interesting to be able to certify advanced verification methods like those involving refinement, alias control, regions, permissions, etc.

An interesting question is how one could certify a VC generator that involves a highly expressive logic, like higher-order logic, as it is the case of the *CFML* method [70] which allows one to use the whole *Coq* language to specify the expected behavior. One challenging aspect of such a certification is that a tool that produces *Coq* definitions, including inductive definitions and module definitions, cannot be directly proved correct in *Coq*, because inductive definitions and module definitions are not first-class objects in *Coq*. Therefore, it seems necessary to involve, in a way or another, a "deep embedding", that is, a formalization of *Coq* in *Coq*, possibly by reusing the deep embedding developed by B. Barras [52].

3.4. Certified Numerical Programs

In recent years, we demonstrated our capability towards specifying and proving properties of floating-point programs, properties which are both complex and precise about the behavior of those programs: see the

publications [65], [109], [61], [104], [64], [59], [98], [96] but also the web galleries of certified programs at our Web page [2], the Hisseo project [3], S. Boldo's page [4], and industrial case studies in the U3CAT ANR project. The ability to express such complex properties comes from models developed in *Coq* [5]. The ability to combine proof by reasoning and proof by computation is a key aspect when dealing with floating-point programs. Such a modeling provides a safe basis when dealing with C source code [4]. However, the proofs can get difficult even on short programs, and to achieve them some automation is needed, and obtained by combining SMT solvers and *Gappa* [62], [79], [46][9]. Finally, the precision of the verification is obtained thanks to precise models of floating-point computations, taking into account the peculiarities of the architecture (e.g. x87 80-bit floating-point unit) and also the compiler optimizations [66], [102].

The directions of research concerning floating-point programs that we want to pursue are the following.

3.4.1. *Making Formal Verification of Floating-point Programs Easier*

A first goal is to ease the formal verification of floating-point programs: the primary objective is still to improve the scope and efficiency of our methods, so as to ease further the verification of numerical programs. The on-going development of the Flocq library should be continued towards the formalization of bit-level manipulations and also of exceptional values (e.g. infinities). We believe that good candidates for applications of our techniques are smart algorithms to compute efficiently with floats, which operate at the bit-level. The formalization of real numbers need to be revamped too: higher-level numerical algorithms are usually built on some mathematical properties (e.g. computable approximations of ideal approximations), which then have to be proved during the formal verification of these algorithms.

Easing the verification of numerical programs also implies more automation. SMT solvers are generic provers well-suited for automatically discharging verification conditions, but they tend to be confused by floating-point arithmetic [31]. Our goal is to improve the arithmetic theories of *Alt-Ergo*, so that they support floating-point arithmetic along their other theories, if possible by reusing the heuristics developed for *Gappa*.

3.4.2. *Continuous Quantities, Numerical Analysis*

The goal is to handle floating-point programs that are related to continuous quantities. This includes numerical analysis programs we have already worked on [14] [61][3]. But our work is only a beginning: we were able to solve the difficulties to prove one particular scheme for one particular partial differential equation. We need to be able to easily prove this kind of programs. This requires new results that handle generic schemes and many partial differential equations. The idea is to design a toolbox to prove these programs with as much automation as possible. We wish this could be used by numerical analysts that are not or hardly familiar with formal methods, but are interested in the formal correctness of their schemes and their programs.

Another very interesting kind of programs (especially for industrial developers) are those based on *hybrid* systems, that is where both discrete and continuous quantities are involved. This is a longer term goal than four years, but we may try to go towards this direction. A first problem is to be able to specify hybrid systems: what are they exactly expected to do? Correctness usually means not going into a forbidden state but we may want additional behavioral properties. A second problem is the interface with continuous systems, such as sensors. How can we describe their behavior? Can we be sure that the formal specification fits? We may think about Ariane V where one piece of code was shamelessly reused from Ariane IV. Ensuring that such a reuse is allowed requires to correctly specify the input ranges and bandwidths of physical sensors.

Studying hybrid systems is among the goals of the new ANR project Cafein.

3.4.3. *Certification of Floating-point Analyses*

In coordination with our second theme, another objective is to port the kernel of *Gappa* into either *Coq* or *Why3*, and then extract a certified executable. Rather than verifying the results of the tool *a posteriori* with a proof checker, they would then be certified *a priori*. This would simplify the inner workings of *Gappa*, help to

---

[2]http://toccata.lri.fr/gallery/index.en.html
[3]http://hisseo.saclay.inria.fr/
[4]http://www.lri.fr/~sboldo/research.html

support new features (e.g. linear arithmetic, elementary functions), and make it scale better to larger formulas, since the tool would no longer need to carry certificates along its computations. Overall the tool would then be able to tackle a wider range of verification conditions.

An ultimate goal would be to develop the decision procedure for floating-point computations, for SMT context, that is mentioned in Section refsec:atp, directly as a certified program in *Coq* or *Why3*.

TYPICAL Project-Team

3. Scientific Foundations

3.1. Logical formalisms

A proof system implements a logical formalism in the way a compiler implements a programming language. Similarly, the choice of the formalism is crucial for the success of the proof system. One of the main line of research of the team is to study or invent type theories that are well-adapted to the formalization of mathematics. For instance a crucial property of a proof system is its correctness, hence the importance of the study of the models of the meta-theory of the Coq proof assistant. An other issue is the interoperability of the various proof systems used to formalize mathematics in the world-wide community of users of proof assistants, and the design of a system which could serve as a back-end to front-end implementing various formalisms and proof languages.

3.2. Libraries of formalized mathematics

It is well known that advanced mathematics can play a crucial role in the design and correctness of sophisticated and sometimes critical software. In some cases, using a proof system is the only option to mechanize the correctness of such programs; this can require the formalization of a wide variety of mathematical theories, and a careful design of these formal libraries for them to be maintainable, combinable and reusable. Furthermore, the ability to formalize advanced contemporary mathematics is still a form of ultimate quality tests for proof systems, and also a way to gain visibility. One of our objectives is to make modern and large pieces of mathematics available as usable formal libraries. Recent examples of complex proofs (Four Color Theorem, Kepler conjecture, classification of finite groups, Fermat theorem) challenge the way the mathematical literature is refereed and published. We think that the development of these formal libraries of mathematics may also change the way certain mathematical result become accepted as theorems. Crafting large bodies of formalized mathematics is a challenging task. These libraries obey similar requirements as software : modularity and usability stem from appropriate data-structures, design patterns and corpus of lemmas. But the appropriate methodology leading to the relevant solutions is often far from obvious, and this is where research has to be done and know-how has to be gained. Up to recently, formal developments were seldom collaborative and rarely benefitted from reusable previous work. The maturity of proof assistants is now sufficient to envision a more modern conception of formal software, as required by large scale verification projects like T. Hales' proof of the Kepler conjecture or the Feit-Thompson theorem. Several members of the TypiCal team are committed in such big formalization projects, or in more specific but related side projects.

3.3. Proof search and automated decision procedures

Interactive proof assistants provide a very expressive logical formalism, rich enough to allow extremely precise descriptions of complex objects like the meta theory of a programming language, a model of C compiler, or the proof of the Four Color Theorem. This description includes logical statements of the properties required by the objects of interest but also their formal proofs, checked by the merciless proof-checker of the system, which should be a small hence trusted piece of code. These systems provide the highest formal guarantee, for instance, of the correctness with respect to the mathematical specification of a code.

Proof-search is a central issue in such a formalization of mathematics. It is also a common aspect of automated reasoning and high-level programming paradigms such as Logic programming. However specific applications commonly involve specific logics or theories, like for instance linear arithmetic. Whether or not such a logical framework can express these at all, it is unlikely that its generic proof-search mechanisms can replace the methods that are specific to a logic or theory. Either because this specific domain lies outside the reach of generic proof-search or simply because generic proof-search is less efficient therein than a purpose-made procedure (typically a decision procedure).

But to enlarge the scope where a specific method applies, one can combine both generic proof search mechanisms with specific methods. We hence investigate how to craft formal proof producing decision procedures in the context of an interactive proof assistant. This activity includes understanding the impact of proof-search mechanism (polarization, focusing, etc.), the implementation of efficient connections between domain specific automated decision procedures (SMT solvers, polynomial optimization tools, etc.) with a proof assistant, and the combination of these two aspects in the design a unique logical framework where a generic notion of proof-search could serve each of the above purposes.

COMMANDS Project-Team

3. Scientific Foundations

3.1. Historical aspects

The roots of deterministic optimal control are the "classical" theory of the calculus of variations, illustrated by the work of Newton, Bernoulli, Euler, and Lagrange (whose famous multipliers were introduced in [69]), with improvements due to the "Chicago school", Bliss [44] during the first part of the 20th century, and by the notion of relaxed problem and generalized solution (Young [77]).

Trajectory optimization really started with the spectacular achievement done by Pontryagin's group [75] during the fifties, by stating, for general optimal control problems, nonlocal optimality conditions generalizing those of Weierstrass. This motivated the application to many industrial problems (see the classical books by Bryson and Ho [50], Leitmann [71], Lee and Markus [70], Ioffe and Tihomirov [66]). Since then, various theoretical achievements have been obtained by extending the results to nonsmooth problems, see Aubin [40], Clarke [51], Ekeland [58].

Dynamic programming was introduced and systematically studied by R. Bellman during the fifties. The HJB equation, whose solution is the value function of the (parameterized) optimal control problem, is a variant of the classical Hamilton-Jacobi equation of mechanics for the case of dynamics parameterized by a control variable. It may be viewed as a differential form of the dynamic programming principle. This nonlinear first-order PDE appears to be well-posed in the framework of *viscosity solutions* introduced by Crandall and Lions [53], [54], [52]. These tools also allow to perform the numerical analysis of discretization schemes. The theoretical contributions in this direction did not cease growing, see the books by Barles [42] and Bardi and Capuzzo-Dolcetta [41].

3.2. Trajectory optimization

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

In the *full discretization approach*, general Runge-Kutta schemes with different values of control for each inner step are used. This allows to obtain and control high orders of precision, see Hager [62], Bonnans [47]. In an interior-point algorithm context, controls can be eliminated and the resulting system of equation is easily solved due to its band structure. Discretization errors due to constraints are discussed in Dontchev et al. [57]. See also Malanowski et al. [72].

In the *indirect* approach, the control is eliminated thanks to Pontryagin's maximum principle. One has then to solve the two-points boundary value problem (with differential variables state and costate) by a single or multiple shooting method. The questions are here the choice of a discretization scheme for the integration of the boundary value problem, of a (possibly globalized) Newton type algorithm for solving the resulting finite dimensional problem in IR^n (n is the number of state variables), and a methodology for finding an initial point.

For state constrained problems or singular arcs, the formulation of the shooting function may be quite elaborate [45], [46], [39]. As initiated in [61], we focus more specifically on the handling of discontinuities, with ongoing work on the geometric integration aspects (Hamiltonian conservation).

3.3. Hamilton-Jacobi-Bellman approach

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

Characterization of the value function From the dynamic programming principle, we derive a characterization of the value function as being a solution (in viscosity sense) of an Hamilton-Jacobi-Bellman equation, which is a nonlinear PDE of dimension equal to the number n of state variables. Since the pioneer works of Crandall and Lions [53], [54], [52], many theoretical contributions were carried out, allowing an understanding of the properties of the value function as well as of the set of admissible trajectories. However, there remains an important effort to provide for the development of effective and adapted numerical tools, mainly because of numerical complexity (complexity is exponential with respect to n).

Numerical approximation for continuous value function Several numerical schemes have been already studied to treat the case when the solution of the HJB equation (the value function) is continuous. Let us quote for example the Semi-Lagrangian methods [60], [59] studied by the team of M. Falcone (La Sapienza, Rome), the high order schemes WENO, ENO, Discrete galerkin introduced by S. Osher, C.-W. Shu, E. Harten [63], [64], [65], [73], and also the schemes on nonregular grids by R. Abgrall [38], [37]. All these schemes rely on finite differences or/and interpolation techniques which lead to numerical diffusions. Hence, the numerical solution is unsatisfying for long time approximations even in the continuous case.

One of the (nonmonotone) schemes for solving the HJB equation is based on the Ultrabee algorithm proposed, in the case of advection equation with constant velocity, by Roe [76] and recently revisited by Després-Lagoutière [56], [55]. The numerical results on several academic problems show the relevance of the antidiffusive schemes. However, the theoretical study of the convergence is a difficult question and is only partially done.

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

Nonsmoothness of the value function. Sometimes the value function is smooth (e.g. in the case of Merton's portfolio problem, Oksendal [78]) and the associated HJB equation can be solved explicitly. Still, the value function is not smooth enough to satisfy the HJB equation in the classical sense. As for the deterministic case, the notion of viscosity solution provides a convenient framework for dealing with the lack of smoothness, see Pham [74], that happens also to be well adapted to the study of discretization errors for numerical discretization schemes [67], [43].

Numerical approximation for optimal stochastic control problems. The numerical discretization of second order HJB equations was the subject of several contributions. The book of Kushner-Dupuis [68] gives a complete synthesis on the Markov chain schemes (i.e Finite Differences, semi-Lagrangian, Finite Elements, ...). Here a main difficulty of these equations comes from the fact that the second order operator (i.e. the diffusion term) is not uniformly elliptic and can be degenerated. Moreover, the diffusion term (covariance matrix) may change direction at any space point and at any time (this matrix is associated the dynamics volatility).

For solving stochastic control problems, we studied the so-called Generalized Finite Differences (GFD), that allow to choose at any node, the stencil approximating the diffusion matrix up to a certain threshold [49]. Determining the stencil and the associated coefficients boils down to a quadratic program to be solved at each point of the grid, and for each control. This is definitely expensive, with the exception of special structures where the coefficients can be computed at low cost. For two dimensional systems, we designed a (very) fast algorithm for computing the coefficients of the GFD scheme, based on the Stern-Brocot tree [48].

DEFI Project-Team

3. Scientific Foundations

3.1. Scientific Foundations

The research activity of our team is dedicated to the design, analysis and implementation of efficient numerical methods to solve inverse and shape/topological optimization problems in connection with wave imaging, structural design, non-destructive testing and medical imaging modalities. We are particularly interested in the development of fast methods that are suited for real-time applications and/or large scale problems. These goals require to work on both the physical and the mathematical models involved and indeed a solid expertise in related numerical algorithms.

This section intends to give a general overview of our research interests and themes. We choose to present them through the specific academic example of inverse scattering problems (from inhomogeneities), which is representative of foreseen developments on both inversion and (tological) optimization methods. The practical problem would be to identify an inclusion from measurements of diffracted waves that result from the interaction of the sought inclusion with some (incident) waves sent into the probed medium. Typical applications include biomedical imaging where using micro-waves one would like to probe the presence of pathological cells, or imaging of urban infrastructures where using ground penetrating radars (GPR) one is interested in finding the location of buried facilities such as pipelines or waste deposits. This kind of applications requires in particular fast and reliable algorithms.

By "imaging" we shall refer to the inverse problem where the concern is only the location and the shape of the inclusion, while "identification" may also indicate getting informations on the inclusion physical parameters.

Both problems (imaging and identification) are non linear and ill-posed (lack of stability with respect to measurements errors if some careful constrains are not added). Moreover, the unique determination of the geometry or the coefficients is not guaranteed in general if sufficient measurements are not available. As an example, in the case of anisotropic inclusions, one can show that an appropriate set of data uniquely determine the geometry but not the material properties.

These theoretical considerations (uniqueness, stability) are not only important in understanding the mathematical properties of the inverse problem, but also guide the choice of appropriate numerical strategies (which information can be stably reconstructed) and also the design of appropriate regularization techniques. Moreover, uniqueness proofs are in general constructive proofs, i.e. they implicitly contain a numerical algorithm to solve the inverse problem, hence their importance for practical applications. The sampling methods introduced below are one example of such algorithms.

A large part of our research activity is dedicated to numerical methods applied to the first type of inverse problems, where only the geometrical information is sought. In its general setting the inverse problem is very challenging and no method can provide a universal satisfactory solution to it (regarding the balance cost-precision-stability). This is why in the majority of the practically employed algorithms, some simplification of the underlying mathematical model is used, according to the specific configuration of the imaging experiment. The most popular ones are geometric optics (the Kirchhoff approximation) for high frequencies and weak scattering (the Born approximation) for small contrasts or small obstacles. They actually give full satisfaction for a wide range of applications as attested by the large success of existing imaging devices (radar, sonar, echography, X-ray tomography, etc.), that rely on one of these approximations.

Generally speaking, the used simplifications result in a linearization of the inverse problem and therefore are usually valid only if the latter is weakly non-linear. The development of these simplified models and the improvement of their efficiency is still a very active research area. With that perspective we are particularly interested in deriving and studying higher order asymptotic models associated with small geometrical parameters such as: small obstacles, thin coatings, wires, periodic media, ... Higher order models usually introduce some non linearity in the inverse problem, but are in principle easier to handle from the numerical point of view than in the case of the exact model.

A larger part of our research activity is dedicated to algorithms that avoid the use of such approximations and that are efficient where classical approaches fail: i.e. roughly speaking when the non linearity of the inverse problem is sufficiently strong. This type of configuration is motivated by the applications mentioned below, and occurs as soon as the geometry of the unknown media generates non negligible multiple scattering effects (multiply-connected and closely spaces obstacles) or when the used frequency is in the so-called resonant region (wave-length comparable to the size of the sought medium). It is therefore much more difficult to deal with and requires new approaches. Our ideas to tackle this problem will be motivated and inspired by recent advances in shape and topological optimization methods and also the introduction of novel classes of imaging algorithms, so-called sampling methods.

The sampling methods are fast imaging solvers adapted to muli-static data (multiple receiver-transmitter pairs) at a fixed frequency. Even if they do not use any linearization the forward model, they rely on computing the solutions to a set of linear problems of small size, that can be performed in a completely parallel procedure. Our team has already a solid expertise in these methods applied to electromagnetic 3-D problems. The success of such approaches was their ability to provide a relatively quick algorithm for solving 3-D problems without any need for a priori knowledge on the physical parameters of the targets. These algorithms solve only the imaging problem, in the sense that only the geometrical information is provided.

Despite the large efforts already spent in the development of this type of methods, either from the algorithmic point of view or the theoretical one, numerous questions are still open. These attractive new algorithms also suffer from the lack of experimental validations, due to their relatively recent introduction. We also would like to invest on this side by developing collaborations with engineering research groups that have experimental facilities. From the practical point of view, the most potential limitation of sampling methods would be the need of a large amount of data to achieve a reasonable accuracy. On the other hand, optimization methods do not suffer from this constrain but they require good initial guess to ensure convergence and reduce the number of iterations. Therefore it seems natural to try to combine the two class of methods in order to calibrate the balance between cost and precision.

Among various shape optimization methods, the Level Set method seems to be particularly suited for such a coupling. First, because it shares similar mechanism as sampling methods: the geometry is captured as a level set of an "indicator function" computed on a caretisian grid. Second, because the two methods do not require any a priori knowledge on the topology of the sought geometry. Beyond the choice of a particular method, the main question would be to define in which way the coupling can be achieved. Obvious strategies consist in using one method to pre-process (initialization) or post-process (find the level set) the other. But one can also think of more elaborate ones, where for instance a sampling method can be used to optimize the choice of the incident wave at each iteration step.The latter point is closely related to the design of so clalled "focusing incident waves" (which are for instance the basis of applications of the time-reversal principle). In the frequency regime, these incident waves can be constructed from the eigenvalue decomposition of the data operator used by sampling methods. The theoretical and numerical investigations of these aspects are still not completely understood for electromagnetic or elastodynamic problems.

Other topological optimization methods, like the homogenization method or the topological gradient method, can also be used, each one provides particular advantages in specific configurations. It is evident that the development of these methods is very suited to inverse problems and provide substantial advantage compared to classical shape optimization methods based on boundary variation. Their applications to inverse problems has not been fully investigated. The efficiency of these optimization methods can also be increased for adequate asymptotic configurations. For instance small amplitude homogenization method can be used as an efficient relaxation method for the inverse problem in the presence of small contrasts. On the other hand, the topological gradient method has shown to perform well in localizing small inclusions with only one iteration.

A broader perspective would be the extension of the above mentioned techniques to time-dependent cases. Taking into account data in time domain is important for many practical applications, such as imaging in cluttered media, the design of absorbing coatings or also crash worthiness in the case of structural design.

For the identification problem, one would like to also have information of the physical properties of the targets. Of course optimization methods is a tool of choice for these problems. However, in some applications only

a qualitative information is needed and obtaining it in a cheaper way can be performed using asymptotic theories combined with sampling methods. We also refer here to the use of so called transmission eigenvalues as qualitative indicators for non destructive testing of dielectrics.

We are also interested in parameter identification problems arising in diffusion-type problems. Our research here is mostly motivated by applications to the imaging of biological tissues with the technique of Diffusion Magnetic Resonance Imaging (DMRI). Roughly speaking DMRI gives a measure of the average distance travelled by water molecules in a certain medium and can give useful information on cellular structure and structural change when the medium is biological tissue. In particular, we would like to infer from DMRI measurements changes in the cellular volume fraction occurring upon various physiological or pathological conditions as well as the average cell size in the case of tumor imaging. The main challenges here are 1) correctly model measured signals using diffusive-type time-dependent PDEs 2) numercially handle the complexity of the tissues 3) use the first two to identify physically relevant parameters from measurements. For the last point we are particularly interested in constructing reduced models of the multiple-compartment Bloch-Torrey partial differential equation using homogenization methods.

DISCO Project-Team

3. Scientific Foundations

3.1. Modeling of complex environment

We want to model phenomena such as a temporary loss of connection (e.g. synchronisation of the movements through haptic interfaces), a nonhomogeneous environment (e.g. case of cryogenic systems) or the presence of the human factor in the control loop (e.g. grid systems) but also problems involved with technological constraints (e.g. range of the sensors). The mathematical models concerned include integro-differential, partial differential equations, algebraic inequalities with the presence of several time scales, whose variables and/or parameters must satisfy certain constraints (for instance, positivity).

3.2. Analysis of interconnected systems

- Algebraic analysis of linear systems

  Study of the structural properties of linear differential time-delay systems and linear infinite-dimensional systems (e.g. invariants, controllability, observability, flatness, reductions, decomposition, decoupling, equivalences) by means of constructive algebra, module theory, homological algebra, algebraic analysis and symbolic computation [8], [9], [85], [106], [87], [90].

- Robust stability of linear systems

  Within an interconnection context, lots of phenomena are modelled directly or after an approximation by delay systems. These systems might have fixed delays, time-varying delays, distributed delays...

  For various infinite-dimensional systems, particularly delay and fractional systems, input-output and time-domain methods are jointly developed in the team to characterize stability. This research is developed at four levels: analytic approaches (H_∞-stability, BIBO-stablity, robust stability, robustness metrics) [1], [2], [5], [6], symbolic computation approaches (SOS methods are used for determining easy-to-check conditions which guarantee that the poles of a given linear system are not in the closed right half-plane, certified CAD techniques), numerical approaches (root-loci, continuation methods) and by means of softwares developed in the team [5], [6].

- Robustness/fragility of biological systems

  Deterministic biological models describing, for instance, species interactions, are frequently composed of equations with important disturbances and poorly known parameters. To evaluate the impact of the uncertainties, we use the techniques of designing of global strict Lyapunov functions or functional developed in the team.

  However, for other biological systems, the notion of robustness may be different and this question is still in its infancy (see, e.g. [98]). Unlike engineering problems where a major issue is to maintain stability in the presence of disturbances, a main issue here is to maintain the system response in the presence of disturbances. For instance, a biological network is required to keep its functioning in case of a failure of one of the nodes in the network. The team, which has a strong expertise in robustness for engineering problems, aims at contributing at the develpment of new robustness metrics in this biological context.

3.3. Stabilization of interconnected systems

- Linear systems: Analytic and algebraic approaches are considered for infinite-dimensional linear systems studied within the input-output framework.

  In the recent years, the Youla-Kučera parametrization (which gives the set of all stabilizing controllers of a system in terms of its coprime factorizations) has been the cornerstone of the success of the H_∞-control since this parametrization allows one to rewrite the problem of finding the optimal stabilizing controllers for a certain norm such as H_∞ or H_2 as affine, and thus, convex problem.

  A central issue studied in the team is the computation of such factorizations for a given infinite-dimensional linear system as well as establishing the links between stabilizability of a system for a certain norm and the existence of coprime factorizations for this system. These questions are fundamental for robust stabilization problems [1], [2], [8], [9].

  We also consider simultaneous stabilization since it plays an important role in the study of reliable stabilization, i.e. in the design of controllers which stabilize a finite family of plants describing a system during normal operating conditions and various failed modes (e.g. loss of sensors or actuators, changes in operating points) [9]. Moreover, we investigate strongly stabilizable systems [9], namely systems which can be stabilized by stable controllers, since they have a good ability to track reference inputs and, in practice, engineers are reluctant to use unstable controllers especially when the system is stable.

- Nonlinear systems

  The project aims at developing robust stabilization theory and methods for important classes of nonlinear systems that ensure good controller performance under uncertainty and time delays. The main techniques include techniques called backstepping and forwarding, contructions of strict Lyapunov functions through so-called "strictification" approaches [3] and construction of Lyapunov-Krasovskii functionals [4], [5], [6].

- Predictive control

  For highly complex systems described in the time-domain and which are submitted to constraints, predictive control seems to be well-adapted. This model based control method (MPC: Model Predictive Control) is founded on the determination of an optimal control sequence over a receding horizon. Due to its formulation in the time-domain, it is an effective tool for handling constraints and uncertainties which can be explicitly taken into account in the synthesis procedure [7]. The team considers how mutiparametric optimization can help to reduce the computational load of this method, allowing its effective use on real world constrained problems.

  The team also investigates stochastic optimization methods such as genetic algorithm, particle swarm optimization or ant colony [10] as they can be used to optimize any criterion and constraint whatever their mathematical structure is. The developed methodologies can be used by non specialists.

3.4. Synthesis of reduced complexity controllers

- PID controllers

  Even though the synthesis of control laws of a given complexity is not a new problem, it is still open, even for finite-dimensional linear systems. Our purpose is to search for good families of "simple" (e.g. low order) controllers for infinite-dimensional dynamical systems. Within our approach, PID candidates are first considered in the team [2], [31].

- Predictive control

  The synthesis of predictive control laws is concerned with the solution of multiparametric optimization problems. Reduced order controller constraints can be viewed as non convex constraints in the synthesis procedure. Such constraints can be taken into account with stochastic algorithms.

Finally, the development of algorithms based on both symbolic computation and numerical methods, and their implementations in dedicated Scilab/Matlab/Maple toolboxes are important issues in the project.

GECO Team

3. Scientific Foundations

3.1. Geometric control theory

The main research topic of the project-team will be **geometric control**, with a special focus on **control design**. The application areas that we target are control of quantum mechanical systems, neurogeometry and switched systems.

Geometric control theory provides a viewpoint and several tools, issued in particular from differential geometry, to tackle typical questions arising in the control framework: controllability, observability, stabilization, optimal control... [27], [62] The geometric control approach is particularly well suited for systems involving nonlinear and nonholonomic phenomena. We recall that nonholonomicity refers to the property of a velocity constraint that is not equivalent to a state constraint.

The expression **control design** refers here to all phases of the construction of a control law, in a mainly open-loop perspective: modeling, controllability analysis, output tracking, motion planning, simultaneous control algorithms, tracking algorithms, performance comparisons for control and tracking algorithms, simulation and implementation.

We recall that

- **controllability** denotes the property of a system for which any two states can be connected by a trajectory corresponding to an admissible control law ;
- **output tracking** refers to a control strategy aiming at keeping the value of some functions of the state arbitrarily close to a prescribed time-dependent profile. A typical example is **configuration tracking** for a mechanical system, in which the controls act as forces and one prescribes the position variables along the trajectory, while the evolution of the momenta is free. One can think for instance at the lateral movement of a car-like vehicle: even if such a movement is unfeasible, it can be tracked with arbitrary precision by applying a suitable control strategy;
- **motion planning** is the expression usually denoting the algorithmic strategy for selecting one control law steering the system from a given initial state to an attainable final one;
- **simultaneous control** concerns algorithms that aim at driving the system from two different initial conditions, with the same control law and over the same time interval, towards two given final states (one can think, for instance, at some control action on a fluid whose goal is to steer simultaneously two floating bodies.) Clearly, the study of which pairs (or n-uples) of states can be simultaneously connected thanks to an admissible control requires an additional controllability analysis with respect to the plain controllability mentioned above.

At the core of control design is then the notion of motion planning. Among the motion planning methods, a preeminent role is played by those based on the Lie algebra associated with the control system ([83], [70], [76]), those exploiting the possible flatness of the system ([56]) and those based on the continuation method ([95]). Optimal control is clearly another method for choosing a control law connecting two states, although it generally introduces new computational and theoretical difficulties.

Control systems with special structure, which are very important for applications are those for which the controls appear linearly. When the controls are not bounded, this means that the admissible velocities form a distribution in the tangent bundle to the state manifold. If the distribution is equipped with a smoothly varying norm (representing a cost of the control), the resulting geometrical structure is called *sub-Riemannian*. Sub-Riemannian geometry thus appears as the underlying geometry of the nonholonomic control systems, playing the same role as Euclidean geometry for linear systems. As such, its study is fundamental for control design. Moreover its importance goes far beyond control theory and is an active field of research both in differential geometry ([82]), geometric measure theory ([57], [31]) and hypoelliptic operator theory ([43]).

Other important classes of control systems are those modeling mechanical systems. The dynamics are naturally defined on the tangent or cotangent bundle of the configuration manifold, they have Lagrangian or Hamiltonian structure, and the controls act as forces. When the controls appear linearly, the resulting model can be seen somehow as a second-order sub-Riemannian structure (see [49]).

The control design topics presented above naturally extend to the case of distributed parameter control systems. The geometric approach to control systems governed by partial differential equations is a novel subject with great potential. It could complement purely analytical and numerical approaches, thanks to its more dynamical, qualitative and intrinsic point of view. An interesting example of this approach is the paper [28] about the controllability of Navier–Stokes equation by low forcing modes.

MAXPLUS Project-Team

3. Scientific Foundations

3.1. L'algèbre max-plus/Max-plus algebra

Le semi-corps *max-plus* est l'ensemble $\mathbb{R} \cup \{-\infty\}$, muni de l'addition $(a, b) \mapsto a \oplus b = \max(a, b)$ et de la multiplication $(a, b) \mapsto a \otimes b = a + b$. Cette structure algébrique diffère des structures de corps classiques par le fait que l'addition n'est pas une loi de groupe, mais est idempotente: $a \oplus a = a$. On rencontre parfois des variantes de cette structure: par exemple, le semi-corps *min-plus* est l'ensemble $\mathbb{R} \cup \{+\infty\}$ muni des lois $a \oplus b = \min(a, b)$ et $a \otimes b = a + b$, et le semi-anneau *tropical* est l'ensemble $\mathbb{N} \cup \{+\infty\}$ munis des mêmes lois. L'on peut se poser la question de généraliser les constructions de l'algèbre et de l'analyse classique, qui reposent pour une bonne part sur des anneaux ou des corps tels que $\mathbb{Z}$ ou $\mathbb{R}$, au cas de semi-anneaux de type max-plus: tel est l'objet de ce qu'on appelle un peu familièrement "l'algèbre max-plus".

Il est impossible ici de donner une vue complète du domaine. Nous nous bornerons à indiquer quelques références bibliographiques. L'intérêt pour les structures de type max-plus est contemporain de la naissance de la théorie des treillis [102]. Depuis, les structures de type max-plus ont été développées indépendamment par plusieurs écoles, en relation avec plusieurs domaines. Les motivations venant de la Recherche Opérationnelle (programmation dynamique, problèmes de plus court chemin, problèmes d'ordonnancement, optimisation discrète) ont été centrales dans le développement du domaine [95], [122], [173], [178], [179]. Les semi-anneaux de type max-plus sont bien sûr reliés aux algèbres de Boole [82]. L'algèbre max-plus apparaît de manière naturelle en contrôle optimal et dans la théorie des équations aux dérivées partielles d'Hamilton-Jacobi [161], [160], [144], [128], [119], [165], [138], [120], [105], [66]. Elle apparaît aussi en analyse asymptotique (asymptotiques de type WKB [143], [144], [128], grandes déviations [159], asymptotiques à température nulle en physique statistique [84]), puisque l'algèbre max-plus apparaît comme limite de l'algèbre usuelle. La théorie des opérateurs linéaires max-plus peut être vue comme faisant partie de la théorie des opérateurs de Perron-Frobenius non-linéaires, ou de la théorie des applications contractantes ou monotones sur les cônes [129], [149], [141], [73], laquelle a de nombreuse motivations, telles l'économie mathématique [146], et la théorie des jeux [162], [56]. Dans la communauté des systèmes à événements discrets, l'algèbre max-plus a été beaucoup étudiée parce qu'elle permet de représenter de manière linéaire les phénomènes de synchronisation, lesquels déterminent le comportement temporel de systèmes de production ou de réseaux, voir [6]. Parmi les développements récents du domaine, on peut citer le calcul des réseaux [83], [133], qui permet de calculer des bornes pire des cas de certaines mesures de qualité de service. En informatique théorique, l'algèbre max-plus (ou plutôt le semi-anneau tropical) a joué un rôle décisif dans la résolution de problèmes de décision en théorie des automates [168], [125], [169], [130], [151]. Notons finalement, pour information, que l'algèbre max-plus est apparue récemment en géométrie algébrique [118], [172], [145], [171] et en théorie des représentations [107], [76], sous les noms de géométrie et combinatoire tropicales.

Nous décrivons maintenant de manière plus détaillée les sujets qui relèvent directement des intérêts du projet, comme la commande optimale, les asymptotiques, et les systèmes à événements discrets.

English version

The *max-plus* semifield is the set $\mathbb{R} \cup \{-\infty\}$, equipped with the addition $(a, b) \mapsto a \oplus b = \max(a, b)$ and the multiplication $(a, b) \mapsto a \otimes b = a + b$. This algebraic structure differs from classical structures, like fields, in that addition is idempotent: $a \oplus a = a$. Several variants have appeared in the literature: for instance, the *min-plus* semifield is the set $\mathbb{R} \cup \{+\infty\}$ equipped with the laws $a \oplus b = \min(a, b)$ and $a \otimes b = a + b$, and the *tropical* semiring is the set $\mathbb{N} \cup \{+\infty\}$ equipped with the same laws. One can ask the question of extending to max-plus type structures the classical constructions and results of algebra and analysis: this is what is often called in a wide sense "max-plus algebra" or "tropical algebra".

It is impossible to give in this short space a fair view of the field. Let us, however, give a few references. The interest in max-plus type structures is contemporaneous with the early developments of lattice theory [102]. Since that time, max-plus structures have been developed independently by several schools, in relation with several fields. Motivations from Operations Research (dynamic programming, shortest path problems, scheduling problems, discrete optimisation) were central in the development of the field [95], [122], [173], [178], [179]. Of course, max-plus type semirings are related to Boolean algebras [82]. Max-plus algebras arises naturally in optimal control and in the theory of Hamilton-Jacobi partial differential equations [161], [160], [144], [128], [119], [165], [138], [120], [105], [66]. It arises in asymptotic analysis (WKB asymptotics [143], [144], [128], large deviation asymptotics [159], or zero temperature asymptotics in statistical physics [84]), since max-plus algebra appears as a limit of the usual algebra. The theory of max-plus linear operators may be thought of as a part of the non-linear Perron-Frobenius theory, or of the theory of nonexpansive or monotone operators on cones [129], [149], [141], [73], a theory with numerous motivations, including mathematical economy [146] and game theory [162], [56]. In the discrete event systems community, max-plus algebra has been much studied since it allows one to represent linearly the synchronisation phenomena which determine the time behaviour of manufacturing systems and networks, see [6]. Recent developments include the network calculus of [83], [133] which allows one to compute worst case bounds for certain measures of quality of service. In theoretical computer science, max-plus algebra (or rather, the tropical semiring) played a key role in the solution of decision problems in automata theory [168], [125], [169], [130], [151]. We finally note for information that max-plus algebra has recently arisen in algebraic geometry [118], [172], [145], [171] and in representation theory [107], [76], under the names of tropical geometry and combinatorics.

We now describe in more details some parts of the subject directly related to our interests, like optimal control, asymptotics, and discrete event systems.

3.2. Algèbre max-plus, programmation dynamique, et commande optimale/Max-plus algebra, dynamic programming, and optimal control

L'exemple le plus simple d'un problème conduisant à une équation min-plus linéaire est le problème classique du plus court chemin. Considérons un graphe dont les nœuds sont numérotés de 1 à n et dont le coût de l'arc allant du nœud i au nœud j est noté $M_{ij} \in \mathbb{R} \cup \{+\infty\}$. Le coût minimal d'un chemin de longueur k, allant de i à j, est donné par la quantité:

$$v_{ij}(k) = \min_{\ell: \, \ell_0 = i, \, \ell_k = j} \sum_{r=0}^{k-1} M_{\ell_r \ell_{r+1}} \ , \tag{1}$$

où le minimum est pris sur tous les chemins $\ell = (\ell_0, ..., \ell_k)$ de longueur k, de nœud initial $\ell_0 = i$ et de nœud final $\ell_k = j$. L'équation classique de la programmation dynamique s'écrit:

$$v_{ij}(k) = \min_{1 \leq s \leq n} \left(M_{is} + v_{sj}(k-1) \right) \ . \tag{2}$$

On reconnaît ainsi une équation linéaire min-plus :

$$v(k) = Mv(k-1) \ , \tag{3}$$

où on note par la concaténation le produit matriciel induit par la structure de l'algèbre min-plus. Le classique *problème de Lagrange* du calcul des variations,

$$v(x, T) = \inf_{X(\cdot), \, X(0) = x} \int_0^T L(X(t), \dot{X}(t)) \mathrm{d}t + \phi(X(T)) \ , \tag{4}$$

où $X(t) \in \mathbb{R}^n$, pour $0 \leq t \leq T$, et $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ est le Lagrangien, peut être vu comme une version continue de (1), ce qui permet de voir l'équation d'Hamilton-Jacobi que vérifie v,

$$v(\cdot, 0) = \phi, \quad \frac{\partial v}{\partial T} + H(x, \frac{\partial v}{\partial x}) = 0, \qquad H(x, p) = \sup_{y \in \mathbb{R}^n} (-p \cdot y - L(x, y)) , \tag{5}$$

comme une équation min-plus linéaire. En particulier, les solutions de (5) vérifient un principe de superposition min-plus: si v et w sont deux solutions, et si $\lambda, \mu \in \mathbb{R}$, $\inf(\lambda + v, \mu + w)$ est encore solution de (5). Ce point de vue, inauguré par Maslov, a conduit au développement de l'école d'Analyse Idempotente (voir [144], [128], [138]).

La présence d'une structure algébrique sous-jacente permet de voir les solutions stationnaires de (2) et (5) comme des vecteurs propres de la matrice M ou du semi-groupe d'évolution de l'équation d'Hamilton-Jacobi. La valeur propre associée fournit le coût moyen par unité de temps (coût ergodique). La représentation des vecteurs propres (voir [161], [173], [95], [121], [89], [72], [6] pour la dimension finie, et [144], [128] pour la dimension infinie) est intimement liée au théorème de l'autoroute qui décrit les trajectoires optimales quand la durée ou la longueur des chemins tend vers l'infini. Pour l'équation d'Hamilton-Jacobi, des résultats reliés sont apparus récemment en théorie d'"Aubry-Mather" [105].

English version

The most elementary example of a problem leading to a min-plus linear equation is the classical shortest path problem. Consider a graph with nodes $1, ..., n$, and let $M_{ij} \in \mathbb{R} \cup \{+\infty\}$ denote the cost of the arc from node i to node j. The minimal cost of a path of a given length, k, from i to j, is given by (1), where the minimum is taken over all paths $\ell = (\ell_0, ..., \ell_k)$ of length k, with initial node $\ell_0 = i$ and final node $\ell_k = j$. The classical dynamic programming equation can be written as in (2). We recognise the min-plus linear equation (3), where concatenation denotes the matrix product induced by the min-plus algebraic structure. The classical *Lagrange problem* of calculus of variations, given by (4) where $X(t) \in \mathbb{R}^n$, for $0 \leq t \leq T$, and $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the Lagrangian, may be thought of as a continuous version of (1), which allows us to see the Hamilton-Jacobi equation (5) satisfied by v, as a min-plus linear equation. In particular, the solutions of (5) satisfy a min-plus superposition principle: if v and w are two solutions, and if $\lambda, \mu \in \mathbb{R}$, then $\inf(\lambda + v, \mu + w)$ is also a solution of (5). This point of view, due to Maslov, led to the developpement of the school of Idempotent Analysis (see [144], [128], [138]).

The underlying algebraic structure allows one to see stationnary solutions of (2) and (5) as eigenvectors of the matrix M or of the evolution semigroup of the Hamilton-Jacobi equation. The associated eigenvalue gives the average cost per time unit (ergodic cost). The representation of eigenvectors (see [161], [173], [121], [89], [95], [72], [6] for the finite dimension case, and [144], [128] for the infinite dimension case) is intimately related to turnpike theorems, which describe optimal trajectories as the horizon, or path length, tends to infinity. For the Hamilton-Jacobi equation, related results have appeared recently in the "Aubry-Mather" theory [105].

3.3. Applications monotones et théorie de Perron-Frobenius non-linéaire, ou l'approche opératorielle du contrôle optimal et des jeux/Monotone maps and non-linear Perron-Frobenius theory, or the operator approach to optimal control and games

On sait depuis le tout début des travaux en décision markovienne que les opérateurs de la programmation dynamique f de problèmes de contrôle optimal ou de jeux (à somme nulle et deux joueurs), avec critère additif, ont les propriétés suivantes :

$$\begin{aligned} \text{monotonie/}monotonicity & & x \leq y & \Rightarrow & f(x) \leq f(y) , \\ \text{contraction/}nonexpansiveness & & \|f(x) - f(y)\|_\infty & \leq & \|x - y\|_\infty . \end{aligned} \tag{6}$$

Ici, l'opérateur f est une application d'un certain espace de fonctions à valeurs réelles dans lui-même, $\leq$ désigne l'ordre partiel usuel, et $\| \cdot \|_\infty$ désigne la norme sup. Dans le cas le plus simple, l'ensemble des états est $\{1, ..., n\}$ et f est une application de $\mathbb{R}^n$ dans lui-même. Les applications monotones qui sont contractantes pour la norme du sup peuvent être vues comme des généralisations non-linéaires des matrices sous-stochastiques. Une sous-classe utile, généralisant les matrices stochastiques, est formée des applications qui sont monotones et commutent avec l'addition d'une constante [94] (celles ci sont parfois appelées fonctions topicales). Les problèmes de programmation dynamique peuvent être traduits en termes d'opérateurs : l'équation de la programmation dynamique d'un problème de commande optimale à horizon fini s'écrit en effet $x(k) = f(x(k-1))$, où $x(k)$ est la fonction valeur en horizon k et $x(0)$ est donné; la fonction valeur y d'un problème à horizon infini (y compris le cas d'un problème d'arrêt optimal) vérifie $y = f(y)$; la fonction valeur z d'un problème avec facteur d'actualisation $0 < \alpha < 1$ vérifie $z = f(\alpha z)$, etc. Ce point de vue abstrait a été très fructueux, voir par exemple [56]. Il permet d'inclure la programmation dynamique dans la perspective plus large de la théorie de Perron-Frobenius non-linéaire, qui, depuis l'extension du théorème de Perron-Frobenius par Krein et Rutman, traite des applications non linéaires sur des cônes vérifiant des conditions de monotonie, de contraction ou d'homogénéité. Les problèmes auxquels on s'intéresse typiquement sont la structure de l'ensemble des points fixes de f, le comportement asymptotique de f^k, en particulier l'existence de la limite de $f^k(x)/k$ lorsque k tends vers l'infini (afin d'obtenir le coût ergodique d'un problème de contrôle optimal ou de jeux), l'asymptotique plus précise de f^k, à une normalisation près (afin d'obtenir le comportement précis de l'itération sur les valeurs), etc. Nous renvoyons le lecteur à [149] pour un panorama. Signalons que dans [111],[7], des algorithmes inspirés de l'algorithme classique d'itérations sur les politiques du contrôle stochastique ont pu être introduits dans le cas des opérateurs monotones contractants généraux, en utilisant des résultats de structure de l'ensemble des points fixes de ces opérateurs. Les applications de la théorie des applications monotones contractantes ne se limitent pas au contrôle optimal et aux jeux. En particulier, on utilise la même classe d'applications dans la modélisation des systèmes à événements discrets, voir le §3.5 ci-dessous, et une classe semblable d'applications en analyse statique de programmes, voir le §4.4 ci-dessous.

English version

Since the very beginning of Markov decision theory, it has been observed that dynamic programming operators f arising in optimal control or (zero-sum, two player) game problems have Properties (6). Here, the operator f is a self-map of a certain space of real valued functions, equipped with the standard ordering $\leq$ and with the sup-norm $\| \cdot \|_\infty$. In the simplest case, the set of states is $\{1, ..., n\}$, and f is a self-map of $\mathbb{R}^n$. Monotone maps that are nonexpansive in the sup norm may be thought of as nonlinear generalisations of substochastic matrices. A useful subclass, which generalises stochastic matrices, consists of those maps which are monotone and commute with the addition of a constant [94] (these maps are sometimes called topical functions). Dynamic programming problems can be translated in operator terms: the dynamic programming equation for a finite horizon problem can be written as $x(k) = f(x(k-1))$, where $x(k)$ is the value function in horizon k and $x(0)$ is given; the value function y of a problem with an infinite horizon (including the case of optimal stopping) satisfies $y = f(y)$; the value function z of a problem with discount factor $0 < \alpha < 1$ satisfies $z = f(\alpha z)$, etc. This abstract point of view has been very fruitful, see for instance [56]. It allows one to put dynamic programming in the wider perspective of nonlinear Perron-Frobenius theory, which, after the extension of the Perron-Frobenius theorem by Krein and Rutman, studies non-linear self-maps of cones, satisfying various monotonicity, nonexpansiveness, and homogeneity conditions. Typical problems of interests are the structure of the fixed point set of f, the asymptotic behaviour of f^k, including the existence of the limit of $f^k(x)/k$ as k tends to infinity (which yields the ergodic cost in control or games problems), the finer asymptotic behaviour of f^k, possibly up to a normalisation (which yields precise results on value iteration), etc. We shall not attempt to survey this theory here, and will only refer the reader to [149] for more background. In [111],[7], algorithms inspired from the classical policy iterations algorithm of stochastic control have been introduced for general monotone nonexpansive operators, using structural results for the fixed point set of these operators. Applications of monotone or nonexpansive maps are not limited to optimal control and game theory. In particular, we also use the same class of maps as models of discrete event dynamics systems,

see §3.5 below, and we shall see in §4.4 that related classes of maps are useful in the static analysis of computer programs.

3.4. Processus de Bellman/Bellman processes

Un autre point de vue sur la commande optimale est la théorie des *processus de Bellman* [160], [97], [96], [66],[1], qui fournit un analogue max-plus de la théorie des probabilités. Cette théorie a été développée à partir de la notion de *mesure idempotente* introduite par Maslov [143]. Elle établit une correspondance entre probabilités et optimisation, dans laquelle les variables aléatoires deviennent des variables de coût (qui permettent de paramétriser les problèmes d'optimisation), la notion d'espérance conditionnelle est remplacée par celle de coût conditionnel (pris sur un ensemble de solutions faisables), la propriété de Markov correspond au principe de la programmation dynamique de Bellman, et la convergence faible à une convergence de type épigraphe. Les théorèmes limites pour les processus de Bellman (loi des grands nombres, théorème de la limite centrale, lois stables) fournissent des résultats asymptotiques en commande optimale. Ces résultats généraux permettent en particulier de comprendre qualitativement les difficultés d'approximation des solutions d'équations d'Hamilton-Jacobi retrouvés en particulier dans le travail de thèse d'Asma Lakhoua [131], [63].

English version

Another point of view on optimal control is the theory of *Bellman processes* [160], [97], [96], [66], [1] which provides a max-plus analogue of probability theory, relying on the theory of *idempotent measures* due to Maslov [143]. This establishes a correspondence between probability and optimisation, in which random variables become cost variables (which allow to parametrise optimisation problems), the notion of conditional expectation is replaced by a notion of conditional cost (taken over a subset of feasible solutions), the Markov property corresponds to the Bellman's dynamic programming principle, and weak convergence corresponds to an epigraph-type convergence. Limit theorems for Bellman processes (law of large numbers, central limit theorems, stable laws) yield asymptotic results in optimal control. Such general results help in particular to understand qualitatively the difficulty of approximation of Hamilton-Jacobi equations found again in particular in the PhD thesis work of Asma Lakhoua [131], [63].

3.5. Systèmes à événements discrets/Discrete event systems

Des systèmes dynamiques max-plus linéaires, de type (2), interviennent aussi, avec une interprétation toute différente, dans la modélisation des systèmes à événements discrets. Dans ce contexte, on associe à chaque tâche répétitive, i, une fonction *compteur*, $v_i : \mathbb{R} \to \mathbb{N}$, telle que $v_i(t)$ compte le nombre cumulé d'occurrences de la tâche i jusqu'à l'instant t. Par exemple, dans un système de production, $v_i(t)$ compte le nombre de pièces d'un certain type produites jusqu'à l'instant t. Dans le cas le plus simple, qui dans le langage des réseaux de Petri, correspond à la sous-classe très étudiée des graphes d'événements temporisés [85], on obtient des équations min-plus linéaires analogues à (2). Cette observation, ou plutôt, l'observation duale faisant intervenir des fonctions dateurs, a été le point de départ [89] de l'approche max-plus des systèmes à événements discrets [6], qui fournit un analogue max-plus de la théorie des systèmes linéaires classiques, incluant les notions de représentation d'état, de stabilité, de séries de transfert, etc. En particulier, les valeurs propres fournissent des mesures de performance telles que le taux de production. Des généralisations non-linéaires, telles que les systèmes dynamiques min-max [150], [124], ont aussi été étudiées. Les systèmes dynamiques max-plus linéaires aléatoires sont particulièrement utiles dans la modélisation des réseaux [71]. Les modèles d'automates à multiplicités max-plus [109], incluant certains versions temporisées des modèles de traces ou de tas de pièces [113], permettent de représenter des phénomènes de concurrence ou de partage de ressources. Les automates à multiplicités max-plus on été très étudiés par ailleurs en informatique théorique [168], [125], [137], [169], [130], [151]. Ils fournissent des modèles particulièrement adaptés à l'analyse de problèmes d'ordonnancement [136].

English version

Dynamical systems of type (2) also arise, with a different interpretation, in the modelling of discrete event systems. In this context, one associates to every repetitive task, i, a counter function, $v_i : \mathbb{R} \to \mathbb{N}$, such that $v_i(t)$ gives the total number of occurrences of task i up to time t. For instance, in a manufacturing system, $v_i(t)$ will count the number of parts of a given type produced up to time t. In the simplest case, which, in the vocabulary of Petri nets, corresponds to the much studied subclass of timed event graphs [85], we get min-plus linear equations similar to (2). This observation, or rather, the dual observation concerning dater functions, was the starting point [89] of the max-plus approach of discrete event systems [6], which provides some analogue of the classical linear control theory, including notions of state space representations, stability, transfer series, etc. In particular, eigenvalues yield performance measures like the throughput. Nonlinear generalisations, like min-max dynamical systems [150], [124], have been particularly studied. Random max-plus linear dynamical systems are particularly useful in the modelling of networks [71]. Max-plus automata models [109], which include some timed version of trace or heaps of pieces models [113], allow to represent phenomena of concurrency or resource sharing. Note that max-plus automata have been much studied in theoretical computer science [168], [125], [137], [169], [130], [151]. Such automata models are particularly adapted to the analysis of scheduling problems [136].

3.6. Algèbre linéaire max-plus/Basic max-plus algebra

Une bonne partie des résultats de l'algèbre max-plus concerne l'étude des systèmes d'équations linéaires. On peut distinguer trois familles d'équations, qui sont traitées par des techniques différentes : 1) Nous avons déjà évoqué dans les sections 3.2 et 3.3 le problème spectral max-plus $Ax = \lambda x$ et ses généralisations. Celui-ci apparaît en contrôle optimal déterministe et dans l'analyse des systèmes à événements discrets. 2) Le problème $Ax = b$ intervient en commande juste-à-temps (dans ce contexte, le vecteur x représente les dates de démarrage des tâches initiales, b représente certaines dates limites, et on se contente souvent de l'inégalité $Ax \leq b$). Le problème $Ax = b$ est intimement lié au problème d'affectation optimale, et plus généralement au problème de transport optimal. Il se traite via la théorie des correspondances de Galois abstraites, ou théorie de la résiduation [102], [78], [173], [178],[6]. Les versions dimension infinie du problème $Ax = b$ sont reliées aux questions d'analyse convexe abstraite [170], [163], [61] et de dualité non convexe. 3) Le problème linéaire général $Ax = Bx$ conduit à des développements combinatoires intéressants (polyèdres max-plus, déterminants max-plus, symétrisation [123], [152],[6]). Le sujet fait l'objet d'un intérêt récemment renouvelé [98].

English version

An important class of results in max-plus algebra concerns the study of max-plus linear equations. One can distinguish three families of equations, which are handled using different techniques: 1) We already mentioned in Sections 3.2 and 3.3 the max-plus spectral problem $Ax = \lambda x$ and its generalisations, which appears in deterministic optimal control and in performance analysis of discrete event systems. 2) The $Ax = b$ problem arises naturally in just in time problems (in this context, the vector x represents the starting times of initial tasks, b represents some deadlines, and one is often content with the inequality $Ax \leq b$). The $Ax = b$ problem is intimately related with optimal assignment, and more generally, with optimal transportation problems. Its theory relies on abstract Galois correspondences, or residuation theory [102], [78], [173], [178],[6]. Infinite dimensional versions of the $Ax = b$ problem are related to questions of abstract convex analysis [170], [163], [61] and nonconvex duality. 3) The general linear system $Ax = Bx$ leads to interesting combinatorial developments (max-plus polyedra, determinants, symmetrisation [123], [152],[6]). The subject has attracted recently a new attention [98].

3.7. Algèbre max-plus et asymptotiques/Using max-plus algebra in asymptotic analysis

Le rôle de l'algèbre min-plus ou max-plus dans les problèmes asymptotiques est évident si l'on écrit

$$e^{-a/\epsilon} + e^{-b/\epsilon} \asymp e^{-\min(a,b)/\epsilon} \ , \qquad e^{-a/\epsilon} \times e^{-b/\epsilon} = e^{-(a+b)/\epsilon} \ , \tag{7}$$

lorsque $\epsilon \to 0^+$. Formellement, l'algèbre min-plus peut être vue comme la limite d'une déformation de l'algèbre classique, en introduisant le semi-anneau $\mathbb{R}_\epsilon$, qui est l'ensemble $\mathbb{R} \cup \{+\infty\}$, muni de l'addition $(a,b) \mapsto -\epsilon \log\left(e^{-a/\epsilon} + e^{-b/\epsilon}\right)$ et de la multiplication $(a,b) \mapsto a+b$. Pour tout $\epsilon > 0$, $\mathbb{R}_\epsilon$ est isomorphe au semi-corps usuel des réels positifs, $(\mathbb{R}_+, +, \times)$, mais pour $\epsilon = 0^+$, $\mathbb{R}_\epsilon$ n'est autre que le semi-anneau min-plus. Cette idée a été introduite par Maslov [143], motivé par l'étude des asymptotiques de type WKB d'équations de Schrödinger. Ce point de vue permet d'utiliser des résultats algébriques pour résoudre des problèmes d'asymptotiques, puisque les équations limites ont souvent un caractère min-plus linéaire.

Cette déformation apparaît classiquement en théorie des grandes déviations à la loi des grands nombres : dans ce contexte, les objets limites sont des mesures idempotentes au sens de Maslov. Voir [1], [159], [62], pour les relations entre l'algèbre max-plus et les grandes déviations, voir aussi [60], [59], [58] pour des applications de ces idées aux perturbations singulières de valeurs propres. La même déformation est à l'origine de nombreux travaux actuels en géométrie tropicale, à la suite de Viro [172].

English version

The role of min-plus algebra in asymptotic problems becomes obvious when writing Equations (7) when $\epsilon \to 0^+$. Formally, min-plus algebra may be thought of as the limit of a deformation of classical algebra, by introducing the semi-field $\mathbb{R}_\epsilon$, which is the set $\mathbb{R} \cup \{+\infty\}$, equipped with the addition $(a,b) \mapsto -\epsilon \log\left(e^{-a/\epsilon} + e^{-b/\epsilon}\right)$ and the multiplication $(a,b) \mapsto a+b$. For all $\epsilon > 0$, $\mathbb{R}_\epsilon$ is isomorphic to the semi-field of usual real positive numbers, $(\mathbb{R}_+, +, \times)$, but for $\epsilon = 0^+$, $\mathbb{R}_\epsilon$ coincides with the min-plus semiring. This idea was introduced by Maslov [143], motivated by the study of WKB-type asymptotics of Schrödinger equations. This point of view allows one to use algebraic results in asymptotics problems, since the limit equations have often some kind of min-plus linear structure.

This deformation appears classically in large deviation theory: in this context, the limiting objects are idempotent measures, in the sense of Maslov. See [1], [159], [62] for the relation between max-plus algebra and large deviations. See also [60], [59], [58] for the application of such ideas to singular perturbation problems for matrix eigenvalues. The same deformation is at the origin of many current works in tropical geometry, in the line initiated by Viro [172].

POEMS Project-Team

3. Scientific Foundations

3.1. Mathematical analysis and simulation of wave propagation

Our activity relies on the existence of mathematical models established by physicists to model the propagation of waves in various situations. The basic ingredient is a partial differential equation (or a system of partial differential equations) of the hyperbolic type that are often (but not always) linear for most of the applications we are interested in. The prototype equation is the wave equation:

$$\frac{\partial^2 u}{\partial t^2} - c^2 \Delta u = 0,$$

which can be directly applied to acoustic waves but which also constitutes a simplified scalar model for other types of waves (This is why the development of new numerical methods often begins by their application to the wave equation). Of course, taking into account more realistic physics will enrich and complexify the basic models (presence of sources, boundary conditions, coupling of models, integro-differential or non linear terms,...)

It is classical to distinguish between two types of problems associated with these models: the time domain problems and the frequency domain (or time harmonic) problems. In the first case, the time is one of the variables of which the unknown solution depends and one has to face an evolution problem. In the second case (which rigorously makes sense only for linear problems), the dependence with respect to time is imposed a priori (via the source term for instance): the solution is supposed to be harmonic in time, proportional to $e^{i\omega t}$, where $\omega > 0$ denotes the pulsation (also commonly, but improperly, called the frequency). Therefore, the time dependence occurs only through this pulsation which is given a priori and plays the rôle of a parameter: the unknown is only a function of space variables. For instance, the wave equation leads to the Helmholtz wave equation (also called the reduced wave equation) :

$$-c^2 \Delta u - \omega^2 u = 0.$$

These two types of problems, although deduced from the same physical modelling, have very different mathematical properties and require the development of adapted numerical methods.

However, there is generally one common feature between the two problems: the existence of a dimension characteristic of the physical phenomenon: the wavelength. Intuitively, this dimension is the length along which the searched solution varies substantially. In the case of the propagation of a wave in an heterogeneous medium, it is necessary to speak of several wavelengths (the wavelength can vary from one medium to another). This quantity has a fundamental influence on the behavior of the solution and its knowledge will have a great influence on the choice of a numerical method.

Nowadays, the numerical techniques for solving the basic academic and industrial problems are well mastered. A lot of companies have at their disposal computational codes whose limits (in particular in terms of accuracy or robustness) are well known. However, the resolution of complex wave propagation problems close to real applications still poses (essentially open) problems which constitute a real challenge for applied mathematicians. A large part of research in mathematics applied to wave propagation problems is oriented towards the following goals:

- the conception of new numerical methods, more and more accurate and high performing.
- the treatment of more and more complex problems (non local models, non linear models, coupled systems, ...)

- the study of specific phenomena or features such as guided waves, resonances,...
- the development of approximate models in various situations,
- imaging techniques and inverse problems related to wave propagation.

3. Scientific Foundations

3.1. Theoretical aspects: probabilistic modeling of irregularity

The modeling of essentially irregular phenomena is an important challenge, with an emphasis on understanding the sources and functions of this irregularity. Probabilistic tools are well-adapted to this task, provided one can design stochastic models for which the regularity can be measured and controlled precisely. Two points deserve special attention:

- first, the study of regularity has to be *local*. Indeed, in most applications, one will want to act on a system based on local temporal or spatial information. For instance, detection of arrhythmias in ECG or of krachs in financial markets should be performed in "real time", or, even better, ahead of time. In this sense, regularity is a *local* indicator of the *local* health of a system.

- Second, although we have used the term "irregularity" in a generic and somewhat vague sense, it seems obvious that, in real-world phenomena, regularity comes in many colors, and a rigorous analysis should distinguish between them. As an example, at least two kinds of irregularities are present in financial logs: the local "roughness" of the records, and the local density and height of jumps. These correspond to two different concepts of regularity (in technical terms, Hölder exponents and local index of stability), and they both contribute a different manner to financial risk.

In view of the above, the *Regularity* team focuses on the design of methods that:

1. define and study precisely various relevant measures of local regularity,

2. allow to build stochastic models versatile enough to mimic the rapid variations of the different kinds of regularities observed in real phenomena,

3. allow to estimate as precisely and rapidly as possible these regularities, so as to alert systems in charge of control.

Our aim is to address the three items above through the design of mathematical tools in the field of probability (and, to a lesser extent, statistics), and to apply these tools to uncertainty management as described in the following section. We note here that we do not intend to address the problem of controlling the phenomena based on regularity, that would naturally constitute an item 4 in the list above. Indeed, while we strongly believe that generic tools may be designed to measure and model regularity, and that these tools may be used to analyze real-world applications, in particular in the field of uncertainty management, it is clear that, when it comes to control, application-specific tools are required, that we do not wish to address.

The research topics of the *Regularity* team can be roughly divided into two strongly interacting axes, corresponding to two complementary ways of studying regularity:

1. developments of tools allowing to characterize, measure and estimate various notions of local regularity, with a particular emphasis on the stochastic frame,

2. definition and fine analysis of stochastic models for which some aspects of local regularity may be prescribed.

These two aspects are detailed in sections 3.2 and 3.3 below.

3.2. Tools for characterizing and measuring regularity

Fractional Dimensions

Although the main focus of our team is on characterizing *local* regularity, on occasions, it is interesting to use a *global* index of regularity. Fractional dimensions provide such an index. In particular, the *regularization dimension*, that was defined in [30], is well adapted to the study stochastic processes, as its definition allows to build robust estimators in an easy way. Since its introduction, regularization dimension has been used by various teams worldwide in many different applications including the characterization of certain stochastic processes, statistical estimation, the study of mammographies or galactograms for breast carcinomas detection, ECG analysis for the study of ventricular arrhythmia, encephalitis diagnosis from EEG, human skin analysis, discrimination between the nature of radioactive contaminations, analysis of porous media textures, well-logs data analysis, agro-alimentary image analysis, road profile analysis, remote sensing, mechanical systems assessment, analysis of video games, ...(see http://regularity.saclay.inria.fr/theory/localregularity/biblioregdim for a list of works using the regularization dimension).

Hölder exponents

The simplest and most popular measures of local regularity are the pointwise and local Hölder exponents. For a stochastic process $\{X(t)\}_{t \in \mathbb{R}}$ whose trajectories are continuous and nowhere differentiable, these are defined, at a point t_0, as the random variables:

$$\alpha_X(t_0, \omega) = \sup \left\{ \alpha : \limsup_{\rho \to 0} \sup_{t,u \in B(t_0,\rho)} \frac{|X_t - X_u|}{\rho^\alpha} < \infty \right\}, \tag{8}$$

and

$$\widetilde{\alpha}_X(t_0, \omega) = \sup \left\{ \alpha : \limsup_{\rho \to 0} \sup_{t,u \in B(t_0,\rho)} \frac{|X_t - X_u|}{\|t - u\|^\alpha} < \infty \right\}. \tag{9}$$

Although these quantities are in general random, we will omit as is customary the dependency in ω and X and write $\alpha(t_0)$ and $\widetilde{\alpha}(t_0)$ instead of $\alpha_X(t_0, \omega)$ and $\widetilde{\alpha}_X(t_0, \omega)$.

The random functions $t \mapsto \alpha_X(t_0, \omega)$ and $t \mapsto \widetilde{\alpha}_X(t_0, \omega)$ are called respectively the pointwise and local Hölder functions of the process X.

The pointwise Hölder exponent is a very versatile tool, in the sense that the set of pointwise Hölder functions of continuous functions is quite large (it coincides with the set of lower limits of sequences of continuous functions [5]). In this sense, the pointwise exponent is often a more precise tool (*i.e.* it varies in a more rapid way) than the local one, since local Hölder functions are always lower semi-continuous. This is why, in particular, it is the exponent that is used as a basis ingredient in multifractal analysis (see section 3.2). For certain classes of stochastic processes, and most notably Gaussian processes, it has the remarkable property that, at each point, it assumes an almost sure value [16]. SRP, mBm, and processes of this kind (see sections 3.3 and 3.3) rely on the sole use of the pointwise Hölder exponent for prescribing the regularity.

However, α_X obviously does not give a complete description of local regularity, even for continuous processes. It is for instance insensitive to "oscillations", contrarily to the local exponent. A simple example in the deterministic frame is provided by the function $x^\gamma \sin(x^{-\beta})$, where γ, β are positive real numbers. This so-called "chirp function" exhibits two kinds of irregularities: the first one, due to the term x^γ is measured by the pointwise Hölder exponent. Indeed, $\alpha(0) = \gamma$. The second one is due to the wild oscillations around 0, to which α is blind. In contrast, the local Hölder exponent at 0 is equal to $\frac{\gamma}{1+\beta}$, and is thus influenced by the oscillatory behaviour.

Another, related, drawback of the pointwise exponent is that it is not stable under integro-differentiation, which sometimes makes its use complicated in applications. Again, the local exponent provides here a useful complement to α, since $\widetilde{\alpha}$ is stable under integro-differentiation.

Both exponents have proved useful in various applications, ranging from image denoising and segmentation to TCP traffic characterization. Applications require precise estimation of these exponents.

Stochastic 2-microlocal analysis

Neither the pointwise nor the local exponents give a complete characterization of the local regularity, and, although their joint use somewhat improves the situation, it is far from yielding the complete picture.

A fuller description of local regularity is provided by the so-called *2-microlocal analysis*, introduced by J.M. Bony [50]. In this frame, regularity at each point is now specified by two indices, which makes the analysis and estimation tasks more difficult. More precisely, a function f is said to belong to the *2-microlocal space* $C_{x_0}^{s,s'}$, where $s + s' > 0, s' < 0$, if and only if its $m = [s + s']$−th order derivative exists around x_0, and if there exists $\delta > 0$, a polynomial P with degree lower than $[s] - m$, and a constant C, such that

$$\left| \frac{\partial^m f(x) - P(x)}{|x - x_0|^{[s]-m}} - \frac{\partial^m f(y) - P(y)}{|y - x_0|^{[s]-m}} \right| \leq C|x - y|^{s+s'-m}(|x - y| + |x - x_0|)^{-s'-[s]+m}$$

for all x, y such that $0 < |x - x_0| < \delta, 0 < |y - x_0| < \delta$. This characterization was obtained in [23], [31]. See [64], [66] for other characterizations and results. These spaces are stable through integro-differentiation, i.e. $f \in C_x^{s,s'}$ if and only if $f' \in C_x^{s-1,s'}$. Knowing to which space f belongs thus allows to predict the evolution of its regularity after derivation, a useful feature if one uses models based on some kind differential equations. A lot of work remains to be done in this area, in order to obtain more general characterizations, to develop robust estimation methods, and to extend the "2-microlocal formalism" : this is a tool allowing to detect which space a function belongs to, from the computation of the Legendre transform of an auxiliary function known as its *2-microlocal spectrum*. This spectrum provide a wealth of information on the local regularity.

In [16], we have laid some foundations for a stochastic version of 2-microlocal analysis. We believe this will provide a fine analysis of the local regularity of random processes in a direction different from the one detailed for instance in [72].We have defined random versions of the 2-microlocal spaces, and given almost sure conditions for continuous processes to belong to such spaces. More precise results have also been obtained for Gaussian processes. A preliminary investigation of the 2-microlocal behaviour of Wiener integrals has been performed.

Multifractal analysis of stochastic processes

A direct use of the local regularity is often fruitful in applications. This is for instance the case in RR analysis or terrain modeling. However, in some situations, it is interesting to supplement or replace it by a more global approach known as *multifractal analysis* (MA). The idea behind MA is to group together all points with same regularity (as measured by the pointwise Hölder exponent) and to measure the "size" of the sets thus obtained [27], [51], [60]. There are mainly two ways to do so, a geometrical and a statistical one.

In the geometrical approach, one defines the *Hausdorff multifractal spectrum* of a process or function X as the function: $\alpha \mapsto f_h(\alpha) = \dim \{t : \alpha_X(t) = \alpha\}$, where $\dim E$ denotes the Hausdorff dimension of the set E. This gives a fine measure-theoretic information, but is often difficult to compute theoretically, and almost impossible to estimate on numerical data.

The statistical path to MA is based on the so-called *large deviation multifractal spectrum*:

$$f_g(\alpha) = \lim_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{\log N_n^\varepsilon(\alpha)}{\log n},$$

where:

$$N_n^\varepsilon(\alpha) = \#\{k : \alpha - \varepsilon \leq \alpha_n^k \leq \alpha + \varepsilon\},$$

and α_n^k is the "coarse grained exponent" corresponding to the interval $I_n^k = \left[\frac{k}{n}, \frac{k+1}{n}\right]$, *i.e.*:

$$\alpha_n^k = \frac{\log |Y_n^k|}{-\log\ n}.$$

Here, Y_n^k is some quantity that measures the variation of X in the interval I_n^k, such as the increment, the oscillation or a wavelet coefficient.

The large deviation spectrum is typically easier to compute and to estimate than the Hausdorff one. In addition, it often gives more relevant information in applications.

Under very mild conditions (*e.g.* for instance, if the support of f_g is bounded, [40]) the concave envelope of f_g can be computed easily from an auxiliary function, called the *Legendre multifractal spectrum*. To do so, one basically interprets the spectrum f_g as a rate function in a large deviation principle (LDP): define, for $q \in \mathbb{R}$,

$$S_n(q) = \sum_{k=0}^{n-1} |Y_n^k|^q, \tag{10}$$

with the convention $0^q := 0$ for all $q \in \mathbb{R}$. Let:

$$\tau(q) = \liminf_{n\to\infty} \frac{\log S_n(q)}{-\log(n)}.$$

The Legendre multifractal spectrum of X is defined as the Legendre transform τ^* of τ:

$$f_l(\alpha) := \tau^*(\alpha) := \inf_{q\in\mathbb{R}}(q\alpha - \tau(q)).$$

To see the relation between f_g and f_l, define the sequence of random variables $Z_n := \log |Y_n^k|$ where the randomness is through a choice of k uniformly in $\{0,...,n-1\}$. Consider the corresponding moment generating functions:

$$c_n(q) := -\frac{\log E_n[\exp(qZ_n)]}{\log(n)}$$

where E_n denotes expectation with respect to P_n, the uniform distribution on $\{0,...,n-1\}$. A version of Gärtner-Ellis theorem ensures that if $\lim c_n(q)$ exists (in which case it equals $1 + \tau(q)$), and is differentiable, then $c^* = f_g - 1$. In this case, one says that the *weak multifractal formalism* holds, *i.e.* $f_g = f_l$. In favorable cases, this also coincides with f_h, a situation referred to as the *strong multifractal formalism*.

Multifractal spectra subsume a lot of information about the distribution of the regularity, that has proved useful in various situations. A most notable example is the strong correlation reported recently in several works between the narrowing of the multifractal spectrum of ECG and certain pathologies of the heart [61], [63]. Let us also mention the multifractality of TCP traffic, that has been both observed experimentally and proved on simplified models of TCP [2], [47].

Another colour in local regularity: jumps

As noted above, apart from Hölder exponents and their generalizations, at least another type of irregularity may sometimes be observed on certain real phenomena: discontinuities, which occur for instance on financial logs and certain biomedical signals. In this frame, it is of interest to supplement Hölder exponents and their extensions with (at least) an additional index that measures the local intensity and size of jumps. This is a topic we intend to pursue in full generality in the near future. So far, we have developed an approach in the particular frame of *multistable processes*. We refer to section 3.3 for more details.

3.3. Stochastic models

The second axis in the theoretical developments of the *Regularity* team aims at defining and studying stochastic processes for which various aspects of the local regularity may be prescribed.

Multifractional Brownian motion

One of the simplest stochastic process for which some kind of control over the Hölder exponents is possible is probably fractional Brownian motion (fBm). This process was defined by Kolmogorov and further studied by Mandelbrot and Van Ness, followed by many authors. The so-called "moving average" definition of fBm reads as follows:

$$Y_t = \int_{-\infty}^{0} \left[(t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \right] . \mathbb{W}(du) + \int_{0}^{t} (t-u)^{H-\frac{1}{2}} . \mathbb{W}(du),$$

where $\mathbb{W}$ denotes the real white noise. The parameter H ranges in $(0,1)$, and it governs the pointwise regularity: indeed, almost surely, at each point, both the local and pointwise Hölder exponents are equal to H.

Although varying H yields processes with different regularity, the fact that the exponents are constant along any single path is often a major drawback for the modeling of real world phenomena. For instance, fBm has often been used for the synthesis natural terrains. This is not satisfactory since it yields images lacking crucial features of real mountains, where some parts are smoother than others, due, for instance, to erosion.

It is possible to generalize fBm to obtain a Gaussian process for which the pointwise Hölder exponent may be tuned at each point: the *multifractional Brownian motion (mBm)* is such an extension, obtained by substituting the constant parameter $H \in (0,1)$ with a *regularity function* $H : \mathbb{R}_+ \to (0,1)$.

mBm was introduced independently by two groups of authors: on the one hand, Peltier and Levy-Vehel [28] defined the mBm $\{X_t; \ t \in \mathbb{R}_+\}$ from the moving average definition of the fractional Brownian motion, and set:

$$X_t = \int_{-\infty}^{0} \left[(t-u)^{H(t)-\frac{1}{2}} - (-u)^{H(t)-\frac{1}{2}} \right] . \mathbb{W}(du) + \int_{0}^{t} (t-u)^{H(t)-\frac{1}{2}} . \mathbb{W}(du),$$

On the other hand, Benassi, Jaffard and Roux [49] defined the mBm from the harmonizable representation of the fBm, *i.e.*:

$$X_t = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{|\xi|^{H(t)+\frac{1}{2}}} . \widehat{\mathbb{W}}(d\xi),$$

where $\widehat{\mathbb{W}}$ denotes the complex white noise.

The Hölder exponents of the mBm are prescribed almost surely: the pointwise Hölder exponent is $\alpha_X(t) = H(t) \wedge \alpha_H(t)$ a.s., and the local Hölder exponent is $\widetilde{\alpha}_X(t) = H(t) \wedge \widetilde{\alpha}_H(t)$ a.s. Consequently, the regularity of the sample paths of the mBm are determined by the function H or by its regularity. The multifractional Brownian motion is our prime example of a stochastic process with prescribed local regularity.

The fact that the local regularity of mBm may be tuned *via* a functional parameter has made it a useful model in various areas such as finance, biomedicine, geophysics, image analysis, ... A large number of studies have been devoted worldwide to its mathematical properties, including in particular its local time. In addition, there is now a rather strong body of work dealing the estimation of its functional parameter, *i.e.* its local regularity. See http://regularity.saclay.inria.fr/theory/stochasticmodels/bibliombm for a partial list of works, applied or theoretical, that deal with mBm.

Self-regulating processes

We have recently introduced another class of stochastic models, inspired by mBm, but where the local regularity, instead of being tuned "exogenously", is a function of the amplitude. In other words, at each point t, the Hölder exponent of the process X verifies almost surely $\alpha_X(t) = g(X(t))$, where g is a fixed deterministic function verifying certain conditions. A process satisfying such an equation is generically termed a *self-regulating process* (SRP). The particular process obtained by adapting adequately mBm is called the self-regulating multifractional process [3]. Another instance is given by modifying the Lévy construction of Brownian motion [42]. The motivation for introducing self-regulating processes is based on the following general fact: in nature, the local regularity of a phenomenon is often related to its amplitude. An intuitive example is provided by natural terrains: in young mountains, regions at higher altitudes are typically more irregular than regions at lower altitudes. We have verified this fact experimentally on several digital elevation models [7]. Other natural phenomena displaying a relation between amplitude and exponent include temperatures records and RR intervals extracted from ECG [38].

To build the SRMP, one starts from a field of fractional Brownian motions $B(t, H)$, where (t, H) span $[0, 1] \times [a, b]$ and $0 < a < b < 1$. For each fixed H, $B(t, H)$ is a fractional Brownian motion with exponent H. Denote:

$$\overline{X}_{\alpha'}^{\beta'} \;\; = \;\; \alpha' + (\beta' - \alpha') \, \tfrac{X - \min_K(X)}{\max_K(X) - \min_K(X)}$$

the affine rescaling between α' and β' of an arbitrary continuous random field over a compact set K. One considers the following (stochastic) operator, defined almost surely:

$$\begin{aligned}
\Lambda_{\alpha', \beta'} \; : \mathcal{C}\left([0, 1], [\alpha, \beta]\right) &\; \rightarrow \; \mathcal{C}\left([0, 1], [\alpha, \beta]\right) \\
Z(.) &\; \mapsto \; \overline{B(., g\left(Z(.)\right))}_{\alpha'}^{\beta'}
\end{aligned}$$

where $\alpha \le \alpha' < \beta' \le \beta$, α and β are two real numbers, and α', β' are random variables adequately chosen. One may show that this operator is contractive with respect to the sup-norm. Its unique fixed point is the SRMP. Additional arguments allow to prove that, indeed, the Hölder exponent at each point is almost surely $g(t)$.

An example of a two dimensional SRMP with function $g(x) = 1 - x^2$ is displayed on figure 1 .

We believe that SRP open a whole new and very promising area of research.

Multistable processes

Non-continuous phenomena are commonly encountered in real-world applications, *e.g.* financial records or EEG traces. For such processes, the information brought by the Hölder exponent must be supplemented by some measure of the density and size of jumps. Stochastic processes with jumps, and in particular Lévy processes, are currently an active area of research.

The simplest class of non-continuous Lévy processes is maybe the one of stable processes [74]. These are mainly characterized by a parameter $\alpha \in (0, 2]$, the *stability index* ($\alpha = 2$ corresponds to the Gaussian case, that we do not consider here). This index measures in some precise sense the intensity of jumps. Paths of stable processes with α close to 2 tend to display "small jumps", while, when α is near 0, their aspect is governed by large ones.

In line with our quest for the characterization and modeling of various notions of local regularity, we have defined *multistable processes*. These are processes which are "locally" stable, but where the stability index α is now a function of time. This allows to model phenomena which, at times, are "almost continuous", and at others display large discontinuities. Such a behaviour is for instance obvious on almost any sufficiently long financial record.

More formally, a multistable process is a process which is, at each time u, tangent to a stable process [59]. Recall that a process Y is said to be tangent at u to the process Y'_u if:
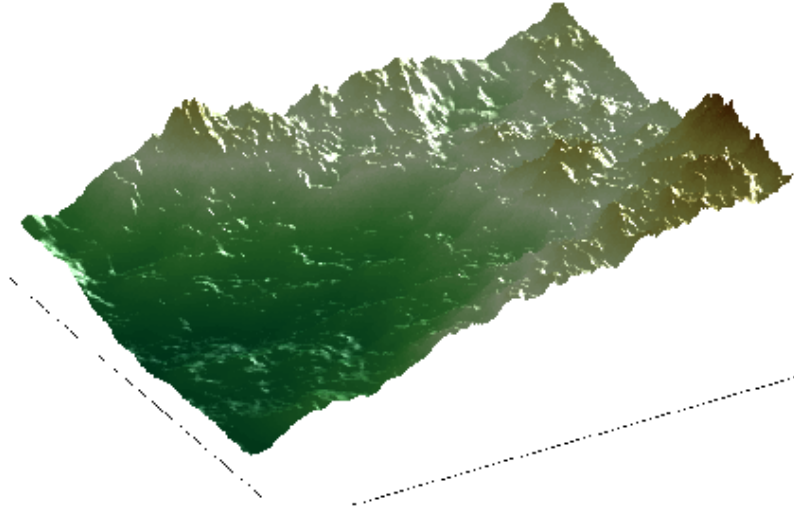
Figure 1. Self-regulating miltifractional process with $g(x) = 1 - x^2$

$$\lim_{r \to 0} \frac{Y(u + rt) - Y(u)}{r^h} = Y_u'(t), \tag{11}$$

where the limit is understood either in finite dimensional distributions or in the stronger sense of distributions. Note Y_u' may and in general will vary with u.

One approach to defining multistable processes is similar to the one developed for constructing mBm [28]: we consider fields of stochastic processes $X(t, u)$, where t is time and u is an independent parameter that controls the variation of α. We then consider a "diagonal" process $Y(t) = X(t, t)$, which will be, under certain conditions, "tangent" at each point t to a process $t \mapsto X(t, u)$.

A particular class of multistable processes, termed "linear multistable multifractional motions" (lmmm) takes the following form [9], [8]. Let $(E, \mathcal{E}, m)$ be a σ-finite measure space, and Π be a Poisson process on $E \times \mathbb{R}$ with mean measure $m \times \mathcal{L}$ ($\mathcal{L}$ denotes the Lebesgue measure). An lmmm is defined as:

$$Y(t) = a(t) \sum_{(\mathsf{X}, \mathsf{Y}) \in \Pi} \mathsf{Y}^{<-1/\alpha(t)>} \left(|t - \mathsf{X}|^{h(t) - 1/\alpha(t)} - |\mathsf{X}|^{h(t) - 1/\alpha(t)} \right) \quad (t \in \mathbb{R}). \tag{12}$$

where $x^{<y>} := \text{sign}(x)|x|^y$, $a : \mathbb{R} \to \mathbb{R}^+$ is a C^1 function and $\alpha : \mathbb{R} \to (0, 2)$ and $h : \mathbb{R} \to (0, 1)$ are C^2 functions.

In fact, lmmm are somewhat more general than said above: indeed, the couple (h, α) allows to prescribe at each point, under certain conditions, both the pointwise Hölder exponent and the local intensity of jumps. In this sense, they generalize both the mBm and the linear multifractional stable motion [75]. From a broader perspective, such multistable multifractional processes are expected to provide relevant models for TCP traces, financial logs, EEG and other phenomena displaying time-varying regularity both in terms of Hölder exponents and discontinuity structure.

Figure 2 displays a graph of an lmmm with linearly increasing α and linearly decreasing H. One sees that the path has large jumps at the beginning, and almost no jumps at the end. Conversely, it is smooth (between jumps) at the beginning, but becomes jaggier and jaggier as time evolves.
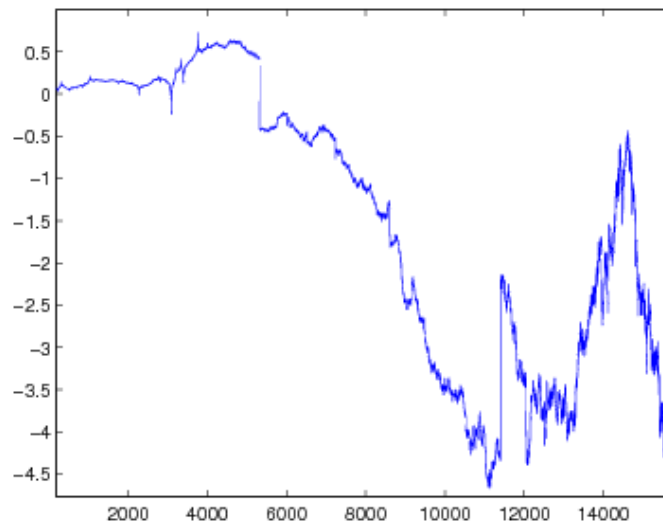

Figure 2. Linear multistable multifractional motion with linearly increasing α and linearly decreasing H

Multiparameter processes

In order to use stochastic processes to represent the variability of multidimensional phenomena, it is necessary to define extensions for indices in $\mathbb{R}^N$ ($N \geq 2$) (see [67] for an introduction to the theory of multiparameter processes). Two different kinds of extensions of multifractional Brownian motion have already been considered: an isotropic extension using the Euclidean norm of $\mathbb{R}^N$ and a tensor product of one-dimensional processes on each axis. We refer to [13] for a comprehensive survey.

These works have highlighted the difficulty of giving satisfactory definitions for increment stationarity, Hölder continuity and covariance structure which are not closely dependent on the structure of $\mathbb{R}^N$. For example, the Euclidean structure can be unadapted to represent natural phenomena.

A promising improvement in the definition of multiparameter extensions is the concept of *set-indexed processes*. A set-indexed process is a process whose indices are no longer "times" or "locations" but may be some compact connected subsets of a metric measure space. In the simplest case, this framework is a generalization of the classical multiparameter processes [62]: usual multiparameter processes are set-indexed processes where the indexing subsets are simply the rectangles $[0, t]$, with $t \in \mathbb{R}_+^N$.

Set-indexed processes allow for greater flexibility, and should in particular be useful for the modeling of censored data. This situation occurs frequently in biology and medicine, since, for instance, data may not be constantly monitored. Censored data also appear in natural terrain modeling when data are acquired from sensors in presence of hidden areas. In these contexts, set-indexed models should constitute a relevant frame.

A set-indexed extension of fBm is the first step toward the modeling of irregular phenomena within this more general frame. In [18], the so-called *set-indexed fractional Brownian motion (sifBm)* was defined as the mean-zero Gaussian process $\{\mathbf{B}_U^H; \ U \in \mathcal{A}\}$ such that

$$\forall U, V \in \mathcal{A}; \quad E[\mathbf{B}_U^H \ \mathbf{B}_V^H] = \frac{1}{2}\left[m(U)^{2H} + m(V)^{2H} - m\left(U \bigtriangleup V\right)^{2H} \right]$$

where $\mathcal{A}$ is a collection of connected compact subsets of a measure metric space and $0 < H \le \frac{1}{2}$.

This process appears to be the only set-indexed process whose projection on increasing paths is a one-parameter fractional Brownian motion [17]. The construction also provides a way to define fBm's extensions on non-euclidean spaces, *e.g.* indices can belong to the unit hyper-sphere of $\mathbb{R}^N$. The study of fractal properties needs specific definitions for increment stationarity and self-similarity of set-indexed processes [20]. We have proved that the sifBm is the only Gaussian set-indexed process satisfying these two (extended) properties.

In the specific case of the indexing collection $\mathcal{A} = \{[0, t], t \in \mathbb{R}_+^N\} \cup \{\varnothing\}$, the sifBm can be seen as a multiparameter extension of fBm which is called *multiparameter fractional Brownian motion (MpfBm)*. This process differs from the Lévy fractional Brownian motion and the fractional Brownian sheet, which are also multiparameter extensions of fBm (but do not derive from set-indexed processes). The local behaviour of the sample paths of the MpfBm has been studied in [12]. The self-similarity index H is proved to be the almost sure value of the local Hölder exponent at any point, and the Hausdorff dimension of the graph is determined in function of H.

The increment stationarity property for set-indexed processes, previously defined in the study of the sifBm, allows to consider set-indexed processes whose increments are independent and stationary. This generalizes the definition of Bass-Pyke and Adler-Feigin for Lévy processes indexed by subsets of $\mathbb{R}^N$, to a more general indexing collection. We have obtained a Lévy-Khintchine representation for these set-indexed Lévy processes and we also characterized this class of Markov processes.

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

3. Scientific Foundations

3.1. General presentation

We learned from the applications we treated that some assumptions which are currently used in asymptotic theory for model selection are often irrelevant in practice. For instance, it is not realistic to assume that the target belongs to the family of models in competition. Moreover, in many situations, it is useful to make the size of the model depend on the sample size which make the asymptotic analysis breakdown. An important aim of SELECT is to propose model selection criteria which take these practical constraints into account.

3.2. A non asymptotic view for model selection

An important purpose of SELECT is to build and analyze penalized log-likelihood model selection criteria that are efficient when the number of models in competition grows to infinity with the number of observations. Concentration inequalities are a key tool for that purpose and lead to data-driven penalty choice strategies. A major issue of SELECT consists of deepening the analysis of data-driven penalties both from the theoretical and the practical side. There is no universal way of calibrating penalties but there are several different general ideas that we want to develop, including heuristics derived from the Gaussian theory, special strategies for variable selection and using resampling methods.

3.3. Taking into account the modeling purpose in model selection

Choosing a model is not only difficult theoretically. From a practical point of view, it is important to design model selection criteria that accommodate situations in which the data probability distribution P is unknown and which take the model user's purpose into account. Most standard model selection criteria assume that P belongs to one of a set of models, without considering the purpose of the model. By also considering the model user's purpose, we avoid or overcome certain theoretical difficulties and can produce flexible model selection criteria with data-driven penalties. The latter is useful in supervised Classification and hidden-structure models.

3.4. Bayesian model selection

The Bayesian approach to statistical problems is fundamentally probabilistic. A joint probability distribution is used to describe the relationships among all the unknowns and the data. Inference is then based on the posterior distribution i.e. the conditional probability distribution of the parameters given the observed data. Exploiting the internal consistency of the probability framework, the posterior distribution extracts the relevant information in the data and provides a complete and coherent summary of post-data uncertainty. Using the posterior to solve specific inference and decision problems is then straightforward, at least in principle.

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

3. Scientific Foundations

3.1. Introduction

This section describes TAO main research directions at the crossroad of Machine Learning and Evolutionary Computation. Since 2008, TAO has been structured in several special interest groups (SIGs) to enable the agile investigation of long-term or emerging theoretical or applicative issues. The comparatively small size of TAO SIGs enables in-depth and lively discussions; the fact that all TAO members belong to several SIGs, on the basis of their personal interests, enforces the strong and informal collaboration of the groups, and the fast information dissemination.

The first two SIGs consolidate the key TAO scientific pillars, while others evolve and adapt to new topics.

The first one, OPT-SIG, addresses stochastic continuous optimization, taking advantage of the fact that TAO is acknowledged the best French research group and one of the top international groups in evolutionary computation from a theoretical and algorithmic standpoint. A main priority on the OPT-SIG research agenda is to provide theoretical and algorithmic guarantees for the current world best continuous stochastic optimizer, CMA-ES, ranging from convergence analysis (Youhei Akimoto's and Verena Heidrich-Meister's post-docs) to a rigorous benchmarking methodology. Incidentally, this benchmark platform has been acknowledged since 2008 as "the" international continuous optimization benchmark, and its extension is at the core of the ANR project NumBBO (starting end 2012). Another priority is to address the current limitations of CMA-ES in terms of high-dimensional or expensive optimization (respectively Ouassim Ait El Hara's and Ilya Loshchilov's PhDs). Mouadh Yagoubi and Zyed Bouzarkouna (resp. CIFRE PSA and IFP-EN) PhD's have continued the EC tradition of industrial breakthrough applications, on which the EC fame solidly relies.

The second SIG, UCT-SIG, benefits from the MoGo expertise and its past and present world records in the domain of computer-Go, establishing the international visibility of TAO in sequential decision making. Since 2010, UCT-SIG resolutely moves to address the problems of energy management from a fundamental and applied perspective. On the one hand, energy management offers a host of challenging issues, ranging from long-horizon policy optimization to the combinatorial nature of the search space, from the modeling of prior knowledge to non-stationary environment to name a few. On the other hand, the energy management issue can hardly be tackled in a pure academic perspective: tight collaborations with industrial partners are needed to access the true operational constraints. Such international and national collaborations have been started by Olivier Teytaud during his one-year stay in Taiwan, and witnessed by the FP7 STREP Citines, the pending ADEME Post contract, and the Ilab with Artelys.

A third SIG, DIS-SIG, is devoted to the modeling and optimization of (large scale) distributed systems. DIS-SIG pursues and extends the goals of the former *Autonomic Computing* SIG, initiated by Cécile Germain and investigating the use of statistical machine learning for large scale computational architectures, from data acquisition (the Grid Observatory in the European Grid Initiative) to grid management and fault detection. More generally, how to model and manage network-based activities has been acknowledged a key topic *per se* in the last months, including: i) the modeling of multi-agent systems and the exploitation of simulation results in the SimTools RNSC network frame; ii) the management of the core communication topology for distributed SAT solving, in the *Microsoft-TAO project* framework. Further extensions are planned in the context of the TIMCO FUI project (started end 2012); the challenge is not only to port ML algorithms on massively distributed architectures, but to see how these architectures can inspire new ML criteria and methodologies.

A fourth SIG, CRI-SIG, focuses on the design of learning and optimization criteria. It elaborates on the lessons learned from the former *Complex Systems* SIG, showing that the key issue in challenging applications is to design the objective itself. Such targeted criteria are pervasive in the study and building of autonomous cognitive systems, ranging from intrinsic rewards in robotics to the notion of saliency in vision and image understanding. The desired criteria can also result from fundamental requirements, such as scale invariance in a statistical physics perspective, and guide the algorithmic design. Additionally, the criteria can also be domain-driven and reflect the expert priors concerning the structure of the sought solution (e.g. spatio-temporal consistency); the challenge is to formulate such criteria in a mixed convex/non differentiable objective function, amenable to tractable optimization.

The activity of the former *Crossing the chasm* SIG gradually decreased after the completion of the 2 PhD theses funded by the Microsoft/Inria joint lab (Adapt project) and devoted to hyper-parameter tuning. Indeed hyper-parameter tuning is still present in TAO, chiefly for continuous optimization (OPT-SIG, section 3.3) and AI planning (CRI-SIG, section 3.4).

3.2. Optimal Decision Making under Uncertainty

Participants: Olivier Teytaud [correspondent], Jean-Joseph Christophe, Adrien Couëtoux, Hassen Doghmen, Jérémie Decock, Nicolas Galichet, Manuel Loth, Marc Schoenauer, Michèle Sebag.

The UCT-SIG works on sequential optimization problems, where a decision has to be made at each time step along a finite time horizon, and the underlying problem involves uncertainties along an either adversarial or stochastic setting.

Application domains include energy management at various time scales and more generally planning, on the one hand, and games (Go, MineSweeper, NoGo [12]) on the other hand.

The main advances done this year include:

- In the domain of computer Go some new performances have been realized [16] and survey papers have been published in Communications of the ACM [10] and as a chapter [65].

- The extension of Upper Confidence Trees to continuous or large domains (states and/or actions) and to domains with high expertise or strong structure has been done [37], [31], [38].

- The extension of Upper Confidence Trees to multi-objective settings has been done, improving on scalarization-based multi-objective reinforcement learning [56].

- Anytime algorithms for discrete time control (note that the classical stochastic dynamic programming is by no means anytime) have been developed and integrated in the Metis software (section 5.1), based on the above cited results.

- Hybrid approaches combining upper confidence trees and e.g. direct policy search or domain specific approaches to yield robust performance w.r.t. long-term effects and take advantage of the combinatorial structure of the domain have been designed, specifically including problem-specific expertise in the playout phase in the domain of job-shop scheduling [53] or MineSweeper [54].

- Optimization algorithms for direct policy search have been designed [55].

- Within the European STREP Mash project, our favorite tools (in particular Monte-Carlo Tree Search) have been extended to difficult settings with no possibility to "undo" a decision [63]. The notion of "risk of exploration" has been investigated [59].

- An experimental analysis of bandit algorithms for small budget cases [36] got the excellent paper award at TAAI 2012.

- In collaboration with Christian Shulte (KTH, Stockholm), one of the main contributors to the well-known general-purpose CP solver *GECODE* (http://www.gecode.org/), and within the Microsoft-Inria joint lab- Adapt project, ideas from UCT have been integrated in GECODE and applied to the job-shop scheduling problem with good first results [52].

- The optimization of low-discrepancy sequences has been done, improving on the best results so far [7]; note that low-discrepancy sequences have been exploited in quite a few of our past works.

3.3. Continuous Optimization

Participants: Ouassim Ait ElHara, Yohei Akimoto, Anne Auger, Zyed Bouzarkouna, Alexandre Chotard, Nikolaus Hansen, Ilya Loshchilov, Verena Heidrich-Meisner, Yann Ollivier, Marc Schoenauer, Michèle Sebag, Olivier Teytaud, Mouadh Yagoubi.

Our main expertise in continuous optimization is on stochastic search algorithms. We address theory, algorithm design and applications. The methods we investigate are adaptive techniques able to learn iteratively parameters of the distribution used to sample solutions. The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is nowadays one of the most powerful methods for derivative-free continuous optimization. We work on different variants of the CMA-ES to improve it in various contexts as described below. We have proven the convergence of simplified variants of the CMA-ES algorithm using the theory of stochastic approximation providing the first proofs of convergence on composite of twice continuously differentiable functions and used Markov chain analysis for analyzing the step-size adaptation rule of the CMA-ES algorithm.

New algorithms based on surrogates and for constrained optimization.   A new variant of CMA-ES to address constraint optimization has been designed [22]. In the Zyed Bouzarkouna's PhD, defended in April 2012, a CIFRE PhD in cooperation with IFP-EN (Institut Français du Pétrole - Energies Nouvelles), new variants of CMA-ES coupled with local-meta-models for expensive optimization have been proposed. They have been applied to well placement problem in oil industry [2]. In the context of his PhD thesis (to be defended in January 2013), Ilya Loshchilov has proposed different surrogate variants of CMA-ES based on ranking-SVM that preserve the invariance to monotonic transformation of the CMA-ES algorithm [51]. He has also explored new restart mechanisms for CMA-ES [47].

Benchmarking.   We have continued our effort for improving standards in benchmarking and pursued the development of the COCO (COmparing Continuous Optimizers) platform. We have organized the ACM GECCO 2012 workshop on Black-Box-Optimization Benchmarking [2] and benchmarked different variants of the CMA-ES algorithms [27] [30] [29] [28] [48] [49] [50]. Our new starting ANR project NumBBO, centered on the COCO platform, aims at extending it for large-scale, expensive, constrained and multi-objective optimization.

Theoretical proofs of convergence.   We have defined and analyzed a variant of CMA-ES being able to prove that its covariance matrix converges to the inverse Hessian on convex-quadratic functions [18]. We have analyzed the convergence of continuous time trajectories associated to step-size adaptive Evolution Strategies on monotonic C^2-composite Functions and proved the local convergence towards local minima [19]. This work has been the starting point for analyzing convergence of step-size adaptive ESs using the framework of stochastic approximation (work about to be submitted). In the context of his thesis, Alexandre Chotard has analyzed the step-size adaptation algorithm of CMA-ES on linear function using the theory of Markov Chains [35].

Multi-objective optimization.   Mouadh Yagoubi has completed his PhD [3], a CIFRE cooperation with PSA (Peugeot-Citroen automotive industry). His work addressed the multi-disciplinary multi-objective optimization of a diesel engine, that led to propose some asynchronous parallelization of expensive objective functions (more than 2 days per evaluation for the complete 3D model) [57].

3.4. Designing criteria

Participants: Jamal Atif, Nicolas Bredèche, Cyril Furtlehner, Yoann Isaac, Victorin Martin, Jean-Marc Montanier, Hélène Paugam-Moisy, Marc Schoenauer, Michèle Sebag.

---

[2]see http://coco.gforge.inria.fr/doku.php?id=bbob-2012

This new SIG, rooted on the claim that *What matters is the criterion*, aims at defining new learning or optimization objectives reflecting fundamental properties of the model, the problem or the expert prior knowledge.

A statistical physics perspective.   In the context of the ANR project TRAVESTI (http://travesti.gforge.inria.fr) which ended this year, we have worked to address more specifically the inverse pairwise Markov random field (MRF) model. On one side we have formalized how the Ising model [64] can be used to perform travel time inference in this context (submitted to AISTATS). Extending the ordinary linear response theory in the vicinity of the so called "Bethe reference point" instead of the interaction-free reference point [67], we were able to provide explicit and tractable formulas to the Plefka's expansion and the related natural gradient at that point. These can be used into various possible algorithms for generating approximate solutions to the inverse Ising problem which we do investigate now. In parallel we have proposed also in [67], a method based on the "iterative proportional scaling" to learn both the factor graph and the couplings by selecting links one by one (submitted to AISTATS). In the Gaussian MRF case this can be implemented efficiently due to local transformations of the precision matrix after adding one link, which is unfortunately not possible in the Ising MRF. It is competitive with L_0 based approach in terms of precision and computational cost, while incidentally the L_1 based method potentially cheaper is not working well for this problem. The flexibility of the method offers in addition the possibility to combine it with spectral constraints like walk-summability with belief propagation or/and graph structure constrain to enforce compatibility with BP. With no additional computational cost we get a complete set of good trade-offs between likelihood and compatibility with BP. Concerning the analysis of the belief propagation we establish in [60] some sufficient condition for encoding a set of local marginals into a stable belief propagation fixed point. Following some work of last year concerning the modeling of congestion at the microscopic level we have finalized in [9] the analysis of a new family of queuing processes where the service rate is coupled stochastically to the number of clients leading a large deviation formulation of the fundamental diagram of traffic flow.

Multi-objective AI Planning.   Within the ANR project DESCARWIN (http://descarwin.lri.fr), Mostepha-Redouane Kouadjia worked on the **multi-objective approach** to AI Planning using the Evolutionary Planner *Divide-and-Evolve*, that evolves a sequential decomposition of the problem at hand: each sub-problem is then solved in turn by some embedded classical planner [72]. Even though the embedded planner is single-objective, DaE can nevertheless handle multi-objective problems: Current work includes the implementation of the multi-objective version of DaE, the definition of some benchmark suite, and some first numerical experiments, comparing in particular the results of a full Pareto approach to those of the classical aggregation method. These works resulted in 3 conference papers recently accepted, introducing a tunable benchmark test suite [45], demonstrating that the best quality measure for parameter tuning in this multi-objective framework is the hypervolume, even in the case of the aggregation approach [46], and comparing the evolutionary multi-objective approach with the aggregation method, the only method known to the AI Planning community [44]. The parameter-tuning algorithm designed for DAE, called Learn-and-Optimize, was published in the selection of papers from *Evolution Artificielle* conference [26]. Though originally designed for Evolutionary AI Planning, the method is applicable to domains where instances sharing similar characteristics w.r.t parameter tuning can be grouped in domains.

Image understanding.   Sequential image understanding refers to the decision making paradigm where objects in an image are successively segmented/recognized following a predefined strategy. Such an approach generally raises some issues about the "best" segmentation sequence to follow and/or how to avoid error propagation. Within the new sequential recognition framework proposed in [8], these issues are addressed as the objects to segment/recognize are represented by a model describing the spatial relations between objects. The process is guided by a criterion derived from visual attention, specifically a saliency map, used to optimize the segmentation sequence. Spatial knowledge is also used to ensure the consistency of the results and to allow backtracking on the segmentation order if needed. The proposed approach was applied for the segmentation of internal brain structures

in magnetic resonance images. The results show the relevance of the optimization criteria and the relevance of the backtracking procedure to guarantee good and consistent results. In [70] we propose a method for simultaneously segmenting and recognizing objects in images, based on a structural representation of the scene and on a constraint propagation method. Within the ANR project LOGIMA, our goal is to address sequential object recognition as an abduction process [69]. Similar principles are at the core of Yoann Isaac's PhD (Digiteo Unsupervised Brain project), in collaboration with CEA LIST. The dictionary-learning approach used to decompose the EEG signal is required to comply with the structure of the data (e.g. spatio-temporal continuity; submitted).

Robotic value systems. Within the European SYMBRION IP, a key milestone toward autonomous cognitive agents has been to provide robots with internal or external rewards, yielding an interesting or competent behavior. Firstly, an objective-free setting referred to as open-ended evolution has been investigated [17], [5], where the criterion to be optimized is left implicit in the reproduction process. in the Secondly, preference-based reinforcement learning has been investigated in Riad Akrour's PhD, where the robot demonstrations are assessed by the expert and these assessments are used to learn a model of the expert's expectations. In [21], this work has been extended and combined with active learning to yield state-of-the art performances with few binary feedbacks from the expert. The hormone-based neural net controller first proposed by T. Schmickl et al. has been thoroughly analyzed and simplified in collaboration with Artificial Life Laboratory from Graz [34].

3.5. Distributed systems

Participants: Cécile Germain-Renaud [correspondent], Philippe Caillou, Dawei Feng, Nadjib Lazaar, Michèle Sebag.

The DIS-SIG explores the issues related to modeling and optimizing distributed systems, ranging from very large scale computational grids to multi-agent systems and distributed constraint solvers.

Coping with non-stationarity. Most existing work on modeling the dynamics of grid behavior assumes a steady-state system and concludes to some form of long-range dependence (slowly decaying correlation) in the associated time-series. But the physical (economic and sociological) processes governing the grid behavior dispel the stationarity hypothesis. When the behavior can be modeled as a time series, an appealing class of models is a sequence of stationary processes separated by break points.The optimisation problem for structural break detection is difficult, because of high dimensionality and a complex objective function. Then, evolutionary algorithms are a method of choice. [15] revisits the optimisation strategy in the light of the general advances in evolutionary computation and the specific opportunities for a separable representation. The single-level optimisation problem is decoupled into a bilevel optimisation. The upper level is the problem of finding the optimal number and location of the break points. The lower level optimises the autoregressive models given the number and locations of break points. At the upper level, our optimisation strategy exploits the state-of-the-art CMA-ES (Covariance Matrix Adaptation - Evolutionary Strategy) instead of the relatively straightforward Genetic Algorithm proposed in the classic AutoPARM fitting procedure for non-stationary time series. The associated representation addresses an important shortcoming of : the distance in the chromosomes space better maps to the distance in the model space. Furthermore, the representation becomes scalable. More precisely, it scales linearly with the length of the data set, independently of the cost of the objective function.

Fault management. Isolating users from the inevitable faults in large distributed systems is critical to Quality of Experience. Thus a significant part of the software infrastructure of large scale distributed systems collects information that will be exploited to discover if, where, and when the system is faulty. In the context of end-to-end probing as the class of monitoring techniques, minimizing the number of probes for a given discovery performance target is critical. While detection and diagnosis have the obvious advantage of providing an explanation of the failure, by exhibiting culprits, they strongly rely on a priori knowledge that is not available for massively distributed

systems. Thus [39], [62] formulates the problem of probe selection for fault prediction based on end-to-end probing as a Collaborative Prediction (CP) problem, based on the reasonable assumption of an underlying factorial model. On an extensive experimental dataset from the EGI grid, the combination of the Maximum Margin Matrix Factorization approach to CP and Active Learning shows excellent performance, reducing the number of probes typically by 80% to 90%.

Multi-agent and games.   The main research focus concerning multi-agent systems was on the observation and automatic description of multi-agent based simulations. Whereas usual parameter space exploration systems observe several experiments, the increasing complexity of simulations makes it harder to understand and describe what happens during a single experiment. It is simple to define global indicators to have an overview of the simulation or to follow individual agents, but the most interesting phenomena often occur at an intermediate level, where *groups of agents* are found. The group level is also suited to analyse and display the dynamics of the model. The online and agent-oriented analysis was achieved and its statistical soundness was assessed [6]. In collaboration with the CEA LIST laboratory, we developed a generic tool (SimAnalyer), which can be used online (with NetLogo) of offline (with Logs) to identify, describe, follow [33] and reproduce [58] clusters of agents in a simulation. To select the most interesting clusters and descriptive variables, new activity indicators reflecting the simulation dynamics have been designed [13].

Parallel SAT Solving.   Recent Parallel SAT solvers use the so-called Conflict-Directed Clause Learning to exchange clauses between the different cores. However, when the number of cores increases, systematic clause sharing leads to communication saturation. Nadjib Lazaar's post-doc, funded by the Microsoft-Inria joint lab, investigated how the communication topology can be optimized online using a Multi-Armed Bandit setting [68], with an improvement of circa 10% (in number of problems solved) and over 50% (in computational time) over ManySAT 2.0, on the 2012 SAT and UNSAT problem suite.

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

3. Scientific Foundations

3.1. RNA and protein structures

3.1.1. RNA

Participants: Julie Bernauer, Alain Denise, Rasmus Fonseca, Feng Lou, Yann Ponty, Mireille Régnier, Philippe Rinaudo, Jean-Marc Steyaert.

Common activity with P. Clote (Boston College and Digiteo).

3.1.1.1. From RNA structure to function

We are currently developing a combinatorial approach, based on random generation, to design small and structured RNAs. An application of such a methodology to the Gag-Pol HIV-1 frameshifting site will be carried out with our collaborators at IGM. We hope that, upon capturing the hybridization energy at the design stage, one will be able to gain control over the rate of frameshift and consequently fine-tune the expression of *Gag/Pol*. Our goal is to build these RNA sequences such that their hybridization with existing mRNAs will be favorable to independent folding, and will therefore affect the stability of some secondary structures involved in recoding events. Moreover it has been observed, mainly on bacteria, that some mRNA sequences may adopt alternate folds. Such events are called a conformational switch, or riboswitch. A common feature of recoding events and riboswitches is that some structural element on the mRNA initiates and unusual action of the ribosome, or allows for an alternate fold under some environmental conditions. One challenge is to predict genes that might be subject to riboswitches.

3.1.1.2. Beyond secondary structure

One of our major challenges is to go beyond secondary structure. Over the past decade, few attempts have been made to predict the 3D structure of RNA from sequence only. So far, few groups have taken this leap. Despite the promises shown by their preliminary results, these approaches currently suffer to a limiting scale due to either their high algorithmic complexity or their difficult automation. Using our expertise in algorithmics and modeling, we plan to design original methods, notably within the AMIS-ARN project (ANR BLANC 2008-2012) in collaboration with PRISM at Versailles University and E.Westhof's group at Strasbourg.

1. *Ab initio* modeling: Starting from the predicted RNA secondary structure, we aim to detect *local structural motifs* in it, giving local 3D conformations. We use the resulting partial structure as a flexible scaffold for a multi-scale reconstruction, notably using game theory. We believe the latter paradigm offers a more realistic view of biological processes than global optimization, used by our competitors, and constitutes a real originality of our project.

2. Comparative modeling: we investigate new algorithms for predicting 3D structures by a comparative approach. This involves comparing multiple RNA sequences and structures at a large scale, that is not possible with current algorithms. Successful methods must rely both on new graph algorithms and on biological expertise on sequence-structure relations in RNA molecules.

3.1.1.3. RNA 3D structure evaluation

The biological function of macromolecules such as proteins and nucleic acids relies on their dynamic structural nature and their ability to interact with many different partners. Their function is mainly determined by the structure those molecules adopt as protein and nucleic acids differ from polypeptides and polynucleotides by their spatial organization. This is specially challenging for RNA where structure flexibility is key.

To address those issues, one has to explore the biologically possible spatial configurations of a macromolecule. The two most common techniques currently used in computational structural biology are Molecular Dynamics (MD) and Monte Carlo techniques (MC). Those techniques require the evaluation of a potential or force-field, which for computational biology are often empirical. They mainly consist of a summation of bonded forces associated with chemical bonds, bond angles, and bond dihedrals, and non-bonded forces associated with van der Waals forces and electrostatic charges. Even if there exists implicit solvent models, they are yet not very well performing and still require a lot of computation time.

Our goal, in collaboration with the Levitt lab at Stanford University and H. van den Bedem at the Stanford Synchrotron Radiation Laboratory (Associate Team ITSNAP http://pages.saclay.inria.fr/julie.bernauer/EA_ITSNAP/) is to develop knowledge-based (KB) potentials, based on measurements on known RNA 3D structures and provide sampling for experimental structure fitting and docking conformation generation. KB potential are quick to evaluate during a simulation and can be used without having to explicitly address the solvent problem. They can be developed at various levels of representation: -atom, base, nucleotide, domain- and could allow the modelling of a wide size range: from a hairpin to the whole ribosome. We also intend to combine these knowledge-based potentials with other potentials (hybrid modelling) and template-based techniques, allowing accurate modelling and dynamics study of very large RNA molecules. Such studies are still a challenge. We will also study conformations for experimental data fitting by extension the innovative, robotics-inspired Kino-Geometric Sampler conformational search algorithm for proteins to nucleic acids and to include experimental data. KGS models a protein as a kinematic linkage and additionally considers hydrogen bonds that "close" kinematic cycles. In closed kinematic cycles rotatable bonds can no longer be deformed independently without breaking closure. KGS preserves all kinematic cycles, and thus hydrogen bonds, by sampling in a subspace of conformational space defined by all closure constraints. KGS exhibits a singularly large search radius and optimally reduces the number of free parameters. These unique features enable flexible 'docking' of atomic models in the data while moderating the risk of overfitting at low resolution. The KGS procedure can also accommodate knowledge-based potentials to improve evaluation of putative conformations and their interactions (see below).

3.1.2. PROTEINS

Participants: Jérôme Azé, Julie Bernauer, Adrien Guilhot-Gaudeffroy, Jean-Marc Steyaert.

3.1.2.1. Docking and evolutionary algorithms

As mentioned above, the function of many proteins depends on their interaction with one or many partners. Docking is the study of how molecules interact. Despite the improvements due to structural genomics initiatives, the experimental solving of complex structures remains a difficult problem. The prediction of complexes, *docking*, proceeds in two steps: a configuration generation phase or *exploration* and an evaluation phase or *scoring*. As the verification of a predicted conformation is time consuming and very expensive, it is a real challenge to reduce the time dedicated to the analysis of complexes by the biologists. Various algorithms and techniques have been used to perform exploration and scoring [49]. The recent rounds of the CAPRI challenge show that real progress has been made using new techniques [46], [3]. Our group has strong experience in cutting edge geometric modelling and scoring techniques using machine learning strategies for protein-protein complexes. In a collaboration with A. Poupon, INRA-Tours, a method that sorts the various potential conformations by decreasing probability of being real complexes has been developed. It relies on a ranking function that is learnt by an evolutionary algorithm. The learning data are given by a geometric modelling of each conformation obtained by the docking algorithm proposed by the biologists. Objective tests are needed for such predictive approaches. The *Critical Assessment of Predicted Interaction*, CAPRI, a community wide experiment modelled after CASP was set up in 2001 to achieve this goal (http://www.ebi.ac.uk/msd-srv/capri/). First results achieved for CAPRI'02 suggested that it is possible to find good conformations by using geometric information for complexes. This approach has been followed (see section New results). As this new algorithm will produce a huge amount of conformations, an adaptation of the ranking function learning step is needed to handle them. In the near future, we intend to extend our approach to protein-RNA complexes.

Such as in the protein case, the function of RNA molecules also depends on their interaction with one or many partners. Upon interaction, RNA molecules often undergo large conformation changes. Understanding how these molecules interact with proteins would allow better targeting for therapeutic studies. The CAPRI (Critical Assessment of PRediction of Interactions) challenge1 has shown that classical docking procedures largely fail when large conformation change occurs and when RNA is involved. This is especially true for RNA molecules, whose large-scale dynamics remain often unknown. Modeling RNA conformational changes is made hard by the inherent flexible nature of their structure but also by the electrostatics involved. These are hard to model and often lead to computationally expensive simulations. Even if for small RNA molecules, molecular dynamics can be used, such simulations are hard to extend to larger molecules and protein-RNA complexes.

For many diseases, such as cancer and HIV, microRNA molecules play a very important role regulating gene expression by guiding the RISC. Some miRNAs have been shown to suppress tumors and are thus ideal candidates for the development of therapeutic agents. Even if various computational techniques have been developed to predict miRNA targets, none of these consider the structural aspects of the interactions between components of the RISC and miRNA. We aim to target these problems, in collaboration with the Huang lab at HKUST (PHC Procore)

The combination of Voronoi models at a coarse-grained level and powerful machine learning techniques allows the accurate scoring of protein-protein complexes [12]. Our actual machine learning approach for proteins is a combination of several machine learning approaches (evolutionary algorithm, decision trees, decision rules,...). By adapting these approaches to protein-RNA, we would have a fast and efficient technique for scoring large protein-RNA complexes where conformational changes are involved.

Working with RNA instead of protein introduce many major differences in the machine learning approaches. RNA conformations are often smaller than protein conformations, which has an impact on the values of the descriptors used to describe objects. Due to the size differences between RNA and protein, it is often more difficult to generate (during the modeling stage) conformations closed to the biological solution (near native solution). The machine learning algorithms therefore need to take into account all theses specificities to be able to learn good predictive models from data that are not very close to the real solution.

The acquired knowledge on RNA flexibility, dynamics and the importance of the sequence will be a strong advance in the modeling of protein-RNA interactions we are working on. Il will help the development of scoring functions based on Voronoi models for RNA and provide us with the level of flexibility needed in complex conformational search. We also intend to develop hybrid KB potentials for complexes from hybrid RNA KB data. These could be incorporated in leading-edge flexible docking modeling software such as Rosetta.

3.2. Annotations

3.2.1. *Word counting*

Participants: Alain Denise, Daria Iakovishina, Yann Ponty, Mireille Régnier, Jean-Marc Steyaert.

We aim at enumerating or generating sequences or structures that are *admissible* in the sense that they are likely to possess some given biological property. Team members have a common expertise in enumeration and random generation of combinatorial structures. They have developped computational tools for probability distributions on combinatorial objects, using in particular generating functions and analytic combinatorics. Admissibility criteria can be mainly statistic; they can also rely on the optimisation of some biological parameter, such as an energy function.

The ability to distinguish a significant event from statistical noise is a crucial need in bioinformatics. In a first step, one defines a suitable probabilistic model (null model) that takes into account the relevant biological properties on the structures of interest. A second step is to develop accurate criteria for assessing (or not) their exceptionality. An event observed in biological sequences, is considered as exceptional, and therefore biologically significant, if the probability that it occurs is very small in the null model. Our approach to

compute such a probability consists in an enumeration of good structures or combinatorial objects. Thirdly, it is necessary to design and implement efficient algorithms to compute these formulae or to generate random data sets. Typical examples that motivate research on words and motifs counting are *Transcription Factor Binding Sites*, TFBSs, consensus models of recoding events and some RNA structural motifs. When relevant motifs do not resort to regular languages, one may still take advantage of combinatorial properties to define functions whose study are amenable to our algebraic tools. One may cite secondary structures and recoding events.

Fast development of high throughput technologies has generated a new challenge for computational biology. The main bottlenecks in applications are the computational analysis of experimental data.

As a first example, numerous new assembling algorithms have recently appeared. Still, the comparison of the results arising from these different algorithms led to significant differences for a given genome assembly. Clearly, strong constraints from the underlying technologies, leading to different data (size, confidence,...) are one origin of the problems and a deeper interpretation is needed, in order to improve algorithms and confidence in the results. One objective is to develop a model of errors, including a statistical model, that takes into account the quality of data for the different technologies, and their volume. This is the subject of an international collaboration with V. Makeev's lab (IoGene, Moscow) and MAGNOME project-team. Second, Next Generation Sequencing open the way to the study of structuralvariants in the genome, as recently described in [44]. Defining a probabilistic model that takes into account main dependencies -such as the GC content- is a task o D. Iakovishina's thesis, in a starting collaboration with V. Boeva (Curie Institute).

3.2.2. *Random generation*

Participants: Alain Denise, Yann Ponty.

Analytical methods may fail when both sequential and structural constraints of sequences are to be modelled or, more generally, when molecular *structures* such as RNA structures have to be handled. The random generation of combinatorial objects is a natural, alternative, framework to assess the significance of observed phenomena. General and efficient techniques have been developed over the last decades to draw objects uniformly at random from an abstract specification. However, in the context of biological sequences and structures, the uniformity assumption becomes unrealistic, and one has to consider non-uniform distributions in order to derive relevant estimates. Typically, context-free grammars can handle certain kinds of long-range interactions such as base pairings in secondary RNA structures. Stochastic context-free grammars (SCFG's) have long been used to model both structural and statistical properties of genomic sequences, particularly for predicting the structure of sequences or for searching for motifs. They can also be used to generate random sequences. However, they do not allow the user to fix the length of these sequences. We developed algorithms for random structures generation that respect a given probability distribution on their components. Our approach is based on the concept of *weighted* combinatorial classes, in combination with the so-named *recursive* method for generating combinatorial structures. To that purpose, one first translates the (biological) structures into combinatorial classes, using the *symbolic method*, an algebraic framework developed by Flajolet *et al*. Adding weights to the atoms allows one to bias the probabilities towards the desired distribution. The main issue is to develop efficient algorithms for finding the suitable weights. An implementation was given in the `GenRGenS` software http://www.lri.fr/~genrgens/, and a generic optimizer that automatically derives suitable parameters for a given grammar, is currently being developped.

In 2005, a new paradigm appeared in the *ab initio* secondary structure prediction [45]: instead of formulating the problem as a classic optimization, this new approach uses statistical sampling within the space of solutions. Besides giving better, more robust, results, it allows for a fruitful adaptation of tools and algorithms derived in a purely combinatorial setting. Indeed, we have done significant and original progress in this area recently [48], [4], including combinatorial models for structures with pseudoknots. Our aim is to combine this paradigm with a fragment based approach for decomposing structures, such as the cycle decomposition used within F. Major's group [47].

Besides, our work on random generation is also applied in a different fields, namely software testing and model-checking, in a continuing collaboration with the Fortesse group at LRI [10], [21].

3.2.3. *Programmed -1 ribosomal frameshifting*

Participants: Patrick Amar, Jérôme Azé, Alain Denise, Christine Froidevaux, Yann Ponty, Cong Zeng.

During protein synthesis, the ribosome decodes the mRNA by assigning a specific amino acid to each codon or nucleotide triplet. Throughout this process the ribosome moves along the mRNA molecule three nucleotides at a time. However, encounters of specific signals found in mRNA from many viruses lead the ribosome to shift one nucleotide backward thus changing its reading frame. We aim at developing a new computational approach that is able to detect these specific signals in genomic databases in order to better understand the molecular choreography leading to the ribosomal frameshifting, which ultimately will help to rationally design new antiviral drugs. As candidates sequences are expected to be numerous, we aim at developing a ranking method to identify the most relevant sequences. Biological testing of these most promising identified candidates by our collaborators from IGM will help us to refine our computational method.

3.2.4. *Knowledge extraction*

Participants: Jérôme Azé, Jiuqiang Chen, Sarah Cohen-Boulakia, Christine Froidevaux.

Our main goal is to design semi-automatic methods for annotation. A possible approach is to focus on the way we could discover relevant motifs in order to make more precise links between function and motifs sequence. For instance, a commonly accepted hypothesis is that function depends on the order of the motifs present in a genomic sequence. Likewise we must be able to evaluate the quality of the annotation obtained. This necessitates giving an estimate of the reliability of the results. This may use combinatorial tools described above. It includes a rigorous statement of the validity domain of algorithms and knowledge of the results provenance. We are interested in provenance resulting from workflow management systems that are important in scientific applications for managing large-scale experiments and can be useful to calculate functional annotations. A given workflow may be executed many times, generating huge amounts of information about data produced and consumed. Given the growing availability of this information, there is an increasing interest in mining it to understand the difference in results produced by different executions.

3.2.5. *Systems Biology*

Participants: Patrick Amar, Sarah Cohen-Boulakia, Alain Denise, Christine Froidevaux, Loic Paulevé, Sabine Peres, Mireille Régnier, Jean-Marc Steyaert.

Systems Biology involves the systematic study of complex interactions in biological systems using an integrative approach. The goal is to find new emergent properties that may arise from the systemic view in order to understand the wide variety of processes that happen in a biological system. Systems Biology activity can be seen as a cycle composed of theory, computational modelling to propose a hypothesis about a biological process, experimental validation, and use of the experimental results to refine or invalidate the computational model (or even the whole theory).

3.2.5.1. Simulations and behavior analysis for metabolism modeling

A great number of methods have been proposed for the study of the behavior of large biological systems. Two methods have been developed and are in use in the team, depending on the specific problems under study : the first one is based on a discrete and direct simulation of the various interactions between the reactants, while the second one deals with an abstract representation by means of differential equations from which we extract various types of features pertaining to the system.

We investigate on the computational modelling step of the cycle by developing a computer simulation system, HSIM, that mimics the interactions of biomolecules in an environment modelling the membranes and compartments found in real cells. In collaboration with biologists from the AMMIS lab. at Rouen we have used HSIM to show the properties of grouping the enzymes of the phosphotransferase system and the glycolytic pathway into metabolons in *E. coli*. In another collaboration with the SYSDIAG Lab (UMR 3145) at Montpellier, we participate at the CompuBioTic project. This is a Synthetic Biology project in the field of medical diagnosis: its goal is to design a small vesicle containing specific proteins and membrane receptors. These components are chosen in a way that their interactions can sense and report the presence in the environment of molecules involved in human pathologies. We used HSIM to help the design and to test qualitatively and quantitatively this *"biological computer"* before *in vitro*.

Given the set of biochemical reactions which describe a metabolic function (e.g. glycolysis, phospholipids' synthesis, etc.) we translate them into a set of o.d.e's whose general form is most often of the Michaelis-Menten type but whose coefficients are usuall very badly determined. The challenge is therefore to extract information as to the system's behavior while making reasonable asumptions on the ranges of values of the parameters. It is sometimes possible to prove mathematically the global stability, but it is also possible to establish it locally in large subdomains by means of simulations. We have developed a software `Mpas` (Metabolic Pathway Analyser Software) that renders the translation in terms of a systems of o.d.e's automatic; then the simulations are also made easy and almost automatic. Furthermore we have developed a method of systematic analysis of the systems in order to characterize those reactants which determine the possible behaviors: usually they are enzymes whose high or low concentrations force the activation of one of the possible branches of the metabolic pathways. A first set of situations has been validated with a research INSERM-INRA team based in Clermont-Ferrand. In particular we have been able to prove mathematically the decisive influence of the enzyme PEMT on the Choline/Ethylamine cycles (M. Behzadi's thesis, defended in 2011).

3.2.5.2. Comparison of Metabolic Networks

In the context of a national project, we study the interest of *fungi* for biomass transformation. Cellulose, hemi-cellulose and lignin are the main components of plant biomass. Their transformation represent a key energy challenges of the 21st century and should eventually allow the production of high value new compounds, such as wood or liquid biofuels (gas or bioethanol). Among the boring organisms, two groups of fungi differ in how they destroy the wood compounds. Analysing new fungi genomes can allow the discover of new species of high interest for bio-transformation.

For a better understanding of how the fungal enzymes facilitates degradation of plant biomass, we conduct a large-scale analysis of the metabolism of fungi. Machine learning approaches such like hierarchical rules prediction will be studied to find new enzymes allowing the transformation of biomass. The KEGG database contains pathways related to fungi and other species. By analysing these known pathways with rules mining approaches, we would be able to predict new enzymes activities.

3.2.5.3. Signalling networks

AMIB and INRA-BIOS (A. Poupon, Tours) are partners in a two years project ASAM (2011-2012). This project aims to help the understanding of signalling pathways involving G protein-coupled receptors (GPCR) which are excellent targets in pharmacogenomics research. Large amounts of experiments are available in this context while globally interpreting all the experimental data remains a very challenging task for biologists. The aim of ASAM is thus to provide means to semi-automatically construct signalling networks of GPCRS.

<p align="center">**GALEN Team**</p>

3. Scientific Foundations

3.1. Structured coupled low- and high-level visual perception

A general framework for the fundamental problems of image segmentation, object recognition and scene analysis is the interpretation of an image in terms of a set of symbols and relations among them. Abstractly stated, image interpretation amounts to mapping an observed image, X to a set of symbols Y. Of particular interest are the symbols Y^* that *optimally explain the underlying image*, as measured by a scoring function s that aims at distinguishing correct (consistent with human labellings) from incorrect interpretations:

$$Y^* = \mathrm{argmax}_Y s(X, Y) \tag{13}$$

Applying this framework requires (a) identifying which symbols and relations to use (b) learning a scoring function s from training data and (c) optimizing over Y in Eq. 1 .

A driving force behind research in GALEN has been the understanding that these three aspects are tightly coupled. In particular, efficient optimization can be achieved by resorting to sparse image representations that 'shortlist' putative solutions and/or by working with scoring functions that can be efficiently optimized. However, the accuracy of a scoring function is largely affected by the breadth of relationships that it accommodates, as well as the completeness of the employed image representation. Determining the tradeoff between these two requirements is far from obvious and often requires approaches customized to the particular problem setting addressed. Summarizing, even though the three problems outlined above can be addressed in isolation, an integrated end-to-end approach is clearly preferable, both for computational efficiency and performance considerations.

Research in GALEN has therefore dealt with the following problem aspects: first, developing a generic and reliable low-level image representation that can be used transversally across multiple tasks. The use of learning-based techniques has been pursued for boundary detection and symmetry detection in [32], yielding state-of-the-art results, while in [27] trajectory grouping was used to come up with a mid-level representation of spatio-temporal data. Complementary to the detection of geometric structures, we have also explored methods for their description both for image and surface data [17]. We are currently pursuing the formulation of the task in structured prediction terms, which will hopefully allow us to exploit the geometrical interdependencies among symmetry and boundary responses.

Second, we have worked on learning scoring functions for detection with deformable models that can leverage upon the developed low-level representations, while also being amenable to efficient optimization. Building on our earlier work on using boundary and symmetry detector responses to perform groupwise registration within categories we used discriminative learning to train hierarchical object models that rely on shape-based representations; these were successfully applied to the detection of shape-based categories, while we are currently pursuing their integration with appearance-based models.

Third, efficient optimization for deformable models was pursued in [18], where we have developed novel techniques for object detection that employ combinatorial optimization tools (A* and Branch-and-Bound) to tame the combinatorial complexity; in particular our work has a best-case performance that is logarithmic in the number of pixels, while our work in [18] allows us to further accelerate object detection by integrating low-level processing (convolutions) with a bounding-based object detection algorithm. Working on a different approach, in [10] we have pursued the exploitation of reinforcement-learning to optimize over the set of shapes derivable from shape grammars. We are currently pursuing a full-fledged bounding-based inference algorithm, which will integrate the tasks of boundary detection and grouping in a single, integrated object detection algorithm.

3.2. Machine Learning & Structure Prediction

The foundation of statistical inference is to learn a function that minimizes the expected loss of a prediction with respect to some unknown distribution

$$\mathcal{R}(f) = \int \ell(f, x, y) dP(x, y), \tag{14}$$

where $\ell(f, x, y)$ is a problem specific loss function that encodes a penalty for predicting $f(x)$ when the correct prediction is y. In our case, we consider x to be a medical image, and y to be some prediction, e.g. the segmentation of a tumor, or a kinematic model of the skeleton. The loss function, ℓ, is informed by the costs associated with making a specific misprediction. As a concrete example, if the true spatial extent of a tumor is encoded in y, $f(x)$ may make mistakes in classifying healthy tissue as a tumor, and mistakes in classifying diseased tissue as healthy. The loss function should encode the potential physiological damage resulting from erroneously targeting healthy tissue for irradiation, as well as the risk from missing a portion of the tumor.

A key problem is that the distribution P is unknown, and any algorithm that is to estimate f from labeled training examples must additionally make an implicit estimate of P. A central technology of empirical inference is to approximate $\mathcal{R}(f)$ with the empirical risk,

$$\mathcal{R}(f) \approx \widehat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f, x_i, y_i), \tag{15}$$

which makes an implicit assumption that the training samples (x_i, y_i) are drawn i.i.d. from P. Direct minimization of $\widehat{\mathcal{R}}(f)$ leads to overfitting when the function class $f \in \mathcal{F}$ is too rich, and regularization is required:

$$\min_{f \in \mathcal{F}} \lambda \Omega(\|f\|) + \widehat{\mathcal{R}}(f), \tag{16}$$

where Ω is a monotonically increasing function that penalizes complex functions.

Equation (4) is very well studied in classical statistics for the case that the output, $y \in \mathcal{Y}$, is a binary or scalar prediction, but this is not the case in most medical imaging prediction tasks of interest. Instead, complex interdependencies in the output space leads to difficulties in modeling inference as a binary prediction problem. One may attempt to model e.g. tumor segmentation as a series of binary predictions at each voxel in a medical image, but this violates the i.i.d. sampling assumption implicit in Equation (3). Furthermore, we typically gain performance by appropriately modeling the inter-relationships between voxel predictions, e.g. by incorporating pairwise and higher order potentials that encode prior knowledge about the problem domain. It is in this context that we develop statistical methods appropriate to structured prediction in the medical imaging setting.

3.3. Self-Paced Learning with Missing Information

Many tasks in artificial intelligence are solved by building a model whose parameters encode the prior domain knowledge and the likelihood of the observed data. In order to use such models in practice, we need to estimate its parameters automatically using training data. The most prevalent paradigm of parameter estimation is supervised learning, which requires the collection of the inputs x_i and the desired outputs y_i. However, such an approach has two main disadvantages. First, obtaining the ground-truth annotation of high-level applications, such as a tight bounding box around all the objects present in an image, is often expensive. This prohibits the use of a large training dataset, which is essential for learning the existing complex models. Second, in many applications, particularly in the field of medical image analysis, obtaining the ground-truth annotation may not be feasible. For example, even the experts may disagree on the correct segmentation of a microscopical image due to the similarities between the appearance of the foreground and background.

In order to address the deficiencies of supervised learning, researchers have started to focus on the problem of parameter estimation with data that contains hidden variables. The hidden variables model the missing information in the annotations. Obtaining such data is practically more feasible: image-level labels ('contains car','does not contain person') instead of tight bounding boxes; partial segmentation of medical images. Formally, the parameters $\mathbf{w}$ of the model are learned by minimizing the following objective:

$$\min_{\mathbf{w}\in\mathcal{W}} R(\mathbf{w}) + \sum_{i=1}^{n} \Delta(y_i, y_i(\mathbf{w}), h_i(\mathbf{w})). \tag{17}$$

Here, $\mathcal{W}$ represents the space of all parameters, n is the number of training samples, $R(\cdot)$ is a regularization function, and $\Delta(\cdot)$ is a measure of the difference between the ground-truth output y_i and the predicted output and hidden variable pair $(y_i(\mathbf{w}), h_i(\mathbf{w}))$.

Previous attempts at minimizing the above objective function treat all the training samples equally. This is in stark contrast to how a child learns: first focus on easy samples ('learn to add two natural numbers') before moving on to more complex samples ('learn to add two complex numbers'). In our work, we capture this intuition using a novel, iterative algorithm called self-paced learning (SPL). At an iteration t, SPL minimizes the following objective function:

$$\min_{\mathbf{w}\in\mathcal{W},\mathbf{v}\in\{0,1\}^n} R(\mathbf{w}) + \sum_{i=1}^{n} v_i \Delta(y_i, y_i(\mathbf{w}), h_i(\mathbf{w})) - \mu_t \sum_{i=1}^{n} v_i. \tag{18}$$

Here, samples with $v_i = 0$ are discarded during the iteration t, since the corresponding loss is multiplied by 0. The term μ_t is a threshold that governs how many samples are discarded. It is annealed at each iteration, allowing the learner to estimate the parameters using more and more samples, until all samples are used. Our results already demonstrate that SPL estimates accurate parameters for various applications such as image classification, discriminative motif finding, handwritten digit recognition and semantic segmentation. We will investigate the use of SPL to estimate the parameters of the models of medical imaging applications, such as segmentation and registration, that are being developed in the GALEN team. The ability to handle missing information is extremely important in this domain due to the similarities between foreground and background appearances (which results in ambiguities in annotations). We will also develop methods that are capable of minimizing more general loss functions that depend on the (unknown) value of the hidden variables, that is,

$$\min_{\mathbf{w}\in\mathcal{W},\theta\in\Theta} R(\mathbf{w}) + \sum_{i=1}^{n} \sum_{h_i\in\mathcal{H}} \Pr(h_i|x_i, y_i; \theta) \Delta(y_i, h_i, y_i(\mathbf{w}), h_i(\mathbf{w})). \tag{19}$$

Here, θ is the parameter vector of the distribution of the hidden variables h_i given the input x_i and output y_i, and needs to be estimated together with the model parameters $\mathbf{w}$. The use of a more general loss function will allow us to better exploit the freely available data with missing information. For example, consider the case where y_i is a binary indicator for the presence of a type of cell in a microscopical image, and h_i is a tight bounding box around the cell. While the loss function $\Delta(y_i, y_i(\mathbf{w}), h_i(\mathbf{w}))$ can be used to learn to classify an image as containing a particular cell or not, the more general loss function $\Delta(y_i, h_i, y_i(\mathbf{w}), h_i(\mathbf{w}))$ can be used to learn to detect the cell as well (since h_i models its location).

3.4. Discrete Biomedical Image Perception

A wide variety of tasks in medical image analysis can be formulated as discrete labeling problems. In very simple terms, a discrete optimization problem can be stated as follows: we are given a discrete set of variables $\mathcal{V}$, all of which are vertices in a graph $\mathcal{G}$. The edges of this graph (denoted by $\mathcal{E}$) encode the variables' relationships. We are also given as input a discrete set of labels $\mathcal{L}$. We must then assign one label from $\mathcal{L}$ to each variable in $\mathcal{V}$. However, each time we choose to assign a label, say, x_{p_1} to a variable p_1, we are forced to pay a price according to the so-called *singleton* potential function $g_p(x_p)$, while each time we choose to assign a pair of labels, say, x_{p_1} and x_{p_2} to two interrelated variables p_1 and p_2 (two nodes that are connected by an edge in the graph $\mathcal{G}$), we are also forced to pay another price, which is now determined by the so called *pairwise* potential function $f_{p_1 p_2}(x_{p_1}, x_{p_2})$. Both the singleton and pairwise potential functions are problem specific and are thus assumed to be provided as input.

Our goal is then to choose a labeling which will allow us to pay the smallest total price. In other words, based on what we have mentioned above, we want to choose a labeling that minimizes the sum of all the MRF potentials, or equivalently the MRF energy. This amounts to solving the following optimization problem:

$$\arg\min_{\{x_p\}} \mathcal{P}(g, f) = \sum_{p \in \mathcal{V}} g_p(x_p) + \sum_{(p_1, p_2) \in \mathcal{E}} f_{p_1 p_2}(x_{p_1}, x_{p_2}). \tag{20}$$

The use of such a model can describe a number of challenging problems in medical image analysis. However these simplistic models can only account for simple interactions between variables, a rather constrained scenario for high-level medical imaging perception tasks. One can augment the expression power of this model through higher order interactions between variables, or a number of cliques $\{C_i, i \in [1, n] = \{\{p_{i^1}, \cdots, p_{i|C_i|}\}\}$ of order $|C_i|$ that will augment the definition of $\mathcal{V}$ and will introduce hyper-vertices:

$$\arg\min_{\{x_p\}} \mathcal{P}(g, f) = \sum_{p \in \mathcal{V}} g_p(x_p) + \sum_{(p_1, p_2) \in \mathcal{E}} f_{p_1 p_2}(x_{p_1}, x_{p_2}) + \sum_{C_i \in \mathcal{E}} f_{p_1 \cdots p_n}(x_{p_{i^1}}, \cdots, p_{x_{i|C_i|}}). \tag{21}$$

where $f_{p_1 \cdots p_n}$ is the price to pay for associating the labels $(x_{p_{i^1}}, \cdots, p_{x_{i|C_i|}})$ to the nodes $(p_1 \cdots p_{i|C_i|})$. Parameter inference, addressed by minimizing the problem above, is the most critical aspect in computational medicine and efficient optimization algorithms are to be evaluated both in terms of computational complexity as well as of inference performance. State of the art methods include deterministic and non-deterministic annealing, genetic algorithms, max-flow/min-cut techniques and relaxation. These methods offer certain strengths while exhibiting certain limitations, mostly related to the amount of interactions which can be tolerated among neighborhood nodes. In the area of medical imaging where domain knowledge is quite strong, one would expect that such interactions should be enforced at the largest scale possible.

<h1 style="text-align:center; color:red;">MACS Project-Team</h1>

3. Scientific Foundations

3.1. Formulation and analysis of effective and reliable shell elements

Thin structures (beams, plates, shells...) are widely considered in engineering applications. However, most experts agree that the corresponding discretization procedures (finite elements) are not yet sufficiently reliable, in particular as regards shell structures. A major cause of these difficulties lies in the numerical locking phenomena that arise in such formulations [2].

The expertise of the team in this area is internationally well-recognized, both in the mathematical and engineering communities. In particular, we have strongly contributed in analysing – and better explaining – the complex locking phenomena that arise in shell formulations [2]. In addition, we have proposed the first (and only to date) shell finite element procedure that circumvents locking [6]. However, the specific treatment applied to avoid locking in this procedure make it unable to correctly represent membrane-dominated behaviors of structures (namely, when locking is not to be expected). In fact, a "perfect shell element" – namely, with the desired reliability properties mathematically substantiated in a general framework – is still to be discovered, whereas numerous teams work on this issue throughout the world.

Another important (and related) issue that is considered in the team pertains to the design and analysis of numerical procedures that are adapted to industrial applications, i.e. that fulfill some actual industrial specifications. In particular, in the past we have achieved the first mathematical analysis of "general shell elements" – which are based on 3D variational formulations instead of shell models – these elements being among the most widely used and most effective shell elements in engineering practice.

3.2. Stability and control of structures

Stability of structures is – of course – a major concern for designers, in particular to ensure that a structure will not undergo poorly damped (or even unbounded) vibrations. In order to obtain improved stability properties – or to reach nominal specifications with a thinner a lighter design – a control device (whether active, semi-active, or passive) may be used.

The research performed in the team in this area – other than some prospective work on robust control – has been so far primarily focused on the stability of structures interacting with fluid flows. This problem has important applications e.g. in aeronautics (flutter of airplane wings), in civil engineering where the design of long-span bridges is now partly governed by wind effects, and in biomechanics (blood flows in arteries, for instance). Very roughly, the coupling between the structure and the flow can be described as follows: the structural displacements modify the geometry of the fluid domain, hence the fluid flow itself which in turn exerts an action on the structure. The effects of structural displacements on the fluid can be taken into account using ALE techniques, but the corresponding direct simulations are highly CPU-intensive, which makes stability analyses of such coupled problems very costly from a computational point of view. In this context a major objective of our work has been to formulate a simplified model of the fluid-structure interaction problem in order to allow computational assessments of stability at a reasonable cost.

3.3. Modeling and estimation in biomechanics

A keen interest in questions arising from the need to model biomechanical systems – and to discretize such problems – has always been present in the team since its creation. Our work in this field until now has been more specifically focused on the objectives related to our participation in the ICEMA ARC projects and in the CardioSense3D initiative, namely, to formulate a complete continuum mechanics model of a beating heart, and to confront – or "couple", in the terminology of the Inria strategic plan – numerical simulations of the model with actual clinical data via a data assimilation procedure.

Our global approach in this framework thus aims at using measurements of the cardiac activity in order to identify the parameters and state of a global electromechanical heart model, hence to give access to quantities of interest for diagnosing electrical activation and mechanical contraction symptoms. The model we propose is based on a chemically-controlled constitutive law of cardiac myofibre mechanics consistent with the behavior of myosin molecular motors [27]. The resulting sarcomere dynamics is in agreement with the "sliding filament hypothesis" introduced by Huxley. This constitutive law has an electrical quantity as an input which can be independently modeled, considered as given (or measured) data, or as a parameter to be estimated.

PARIETAL Project-Team

3. Scientific Foundations

3.1. Human neuroimaging data and its use

Human neuroimaging consists in acquiring non-invasively image data from normal and diseased human populations. Magnetic Resonance Imaging (MRI) can be used to acquire information on brain structure and function at high spatial resolution.

- T1-weighted MRI is used to obtain a segmentation of the brain into different different tissues, such as gray matter, white matter, deep nuclei, cerebro-spinal fluid, at the millimeter or sub-millimeter resolution. This can then be used to derive geometric and anatomical information on the brain, e.g. cortical thickness.

- Diffusion-weighted MRI measures the local diffusion of water molecules in the brain at the resolution of 2mm, in a set of directions (30 to 60 typically). Local anisotropy, observed in white matter, yields a geometric model of fiber tracts along which water diffusion occurs, and thus provides essential information of the connectivity structure of the brain.

- Functional MRI measures the blood-oxygen-level-dependent (BOLD) contrast that reflects neural activity in the brain, at a spatial resolution of 2 to 3mm, and a temporal resolution of 2-3s. This yields a spatially resolved image of brain functional networks that can be modulated either by specific cognitive tasks or appear as networks of correlated activity.

- Electro- and Magneto-encephalography (MEEG) are two additional modalities that complement functional MRI, as they directly measure the electric and magnetic signals elicited by neural activity, at the millisecond scale. These modalities rely on surface measurements and do not localize brain activity very accurately in the spatial domain.

3.2. High-field MRI

High field MRI as performed at Neurospin (7T on humans, 11.7T in 2013, 17.6T on rats) brings an improvement over traditional MRI acquisitions at 1.5T or 3T, related to to a higher signal-to-noise ratio in the data. Depending on the data and applicative context, this gain in SNR can be traded against spatial resolution improvements, thus helping in getting more detailed views of brain structure and function. This comes at the risk of higher susceptibility distortions of the MRI scans and signal inhomogeneities, that need to be corrected for. Improvements at the acquisition level may come from the use of new coils (such as the new 32 channels coil on the 7T at Neurospin).

3.3. Technical challenges for the analysis of neuroimaging data

The first limitation of Neuroimaging-based brain analysis is the limited Signal-to-Noise Ratio of the data. A particularly striking case if functional MRI, where only a fraction of the data is actually understood, and from which it is impossible to observe by eye the effect of neural activation on the raw data. Moreover, far from traditional i.i.d. Gaussian models, the noise in MRI typically exhibits correlations and long-distance correlation properties (e.g. motion-related signal) and has potentially large amplitude, which can make it hard to distinguish from true signal on a purely statistical basis. A related difficulty is the *lack of salient structure* in the data: it is hard to infer meaningful patterns (either through segmentation or factorization procedures) based on the data only. A typical case is the inference of brain networks from resting-state functional connectivity data.

Regarding statistical methodology, neuroimaging problems also suffer from the relative paucity of the data, i.e. the relatively small number of images available to learn brain features or models, e.g. with respect to the size of the images or the number of potential structures of interest. This leads to several kinds of difficulties, known either as multiple comparison problems or curse of dimensionality. One possibility to overcome this challenge is to increase the amount of data by using images from multiple acquisition centers, at the risk of introducing scanner-related variability, thus challenging the homogeneity of the data. This becomes an important concern with the advent of cross-modal neuroimaging-genetics studies.

<div align="center">**GRAND-LARGE Project-Team**</div>

3. Scientific Foundations

3.1. Large Scale Distributed Systems (LSDS)

What makes a fundamental difference between recent Global Computing systems (Seti@home), Grid (EGEE, TeraGrid) and former works on distributed systems is the large scale of these systems. This characteristic becomes also true for large scale parallel computers gathering tens of thousands of CPU cores. The notion of Large Scale is linked to a set of features that has to be taken into account in these systems. An example is the system dynamicity caused by node volatility: in Internet Computing Platforms (also called Desktop Grids), a non predictable number of nodes may leave the system at any time. Some recent results also report a very low MTTI (Mean Time To Interrupt) in top level supercomputers gathering 100,000+ CPU cores. Another example of characteristics is the complete lack of control of nodes connectivity. In Desktop Grid, we cannot assume that external administrator is able to intervene in the network setting of the nodes, especially their connection to Internet via NAT and Firewalls. This means that we have to deal with the in place infrastructure in terms of performance, heterogeneity, dynamicity and connectivity. These characteristics, associated with the requirement of scalability, establish a new research context in distributed systems. The Grand-Large project aims at investigating theoretically as well as experimentally the fundamental mechanisms of LSDS, especially for the high performance computing applications.

3.1.1. Computing on Large Scale Global Computing systems

Large scale parallel and distributed systems are mainly used in the context of Internet Computing. As a consequence, until Sept. 2007, Grand-Large has focused mainly on Desktop Grids. Desktop Grids are developed for computing (SETI@home, Folding@home, Decrypthon, etc.), file exchanges (Napster, Kazaa, eDonkey, Gnutella, etc.), networking experiments (PlanetLab, Porivo) and communications such as instant messaging and phone over IP (Jabber, Skype). In the High Performance Computing domain, LSDS have emerged while the community was considering clustering and hierarchical designs as good performance-cost tradeoffs. Nowadays, Internet Computing systems are still very popular (the BOINC platform is used to run over 40 Internet Computing projects and XtremWeb is used in production in three countries) and still raise important research issues.

Desktop Grid systems essentially extend the notion of computing beyond the frontier of administration domains. The very first paper discussing this type of systems [88] presented the Worm programs and several key ideas that are currently investigated in autonomous computing (self replication, migration, distributed coordination, etc.). LSDS inherit the principle of aggregating inexpensive, often already in place, resources, from past research in cycle stealing/resource sharing. Due to its high attractiveness, cycle stealing has been studied in many research projects like Condor [77] , Glunix [71] and Mosix [50], to cite a few. A first approach to cross administration domains was proposed by Web Computing projects such as Jet [81], Charlotte [51], Javeline [65], Bayanihan [86], SuperWeb [47], ParaWeb [58] and PopCorn [60]. These projects have emerged with Java, taking benefit of the virtual machine properties: high portability across heterogeneous hardware and OS, large diffusion of virtual machine in Web browsers and a strong security model associated with bytecode execution. Performance and functionality limitations are some of the fundamental motivations of the second generation of Global Computing systems like BOINC [49] and XtremWeb [67]. The second generation of Global Computing systems appeared in the form of generic middleware which allow scientists and programmers to design and set up their own distributed computing project. As a result, we have seen the emergence of large communities of volunteers and projects. Currently, Global Computing systems are among the largest distributed systems in the world. In the mean time, several studies succeeded to understand and enhance the performance of these systems, by characterizing the system resources in term of volatility and heterogeneity and by studying new scheduling heuristics to support new classes of applications: data-intensive, long running application with checkpoint, workflow, soft-real time etc... However, despite these

recent progresses, one can note that Global Computing systems are not yet part of high performance solution, commonly used by scientists. Recent researches to fulfill the requirements of Desktop Grids for high demanding users aim at redesigning Desktop Grid middleware by essentially turning a set of volatile nodes into a virtual cluster and allowing the deployment of regular HPC utilities (batch schedulers, parallel communication libraries, checkpoint services, etc...) on top of this virtual cluster. The new generation would permit a better integration in the environment of the scientists such as computational Grids, and consequently, would broaden the usage of Desktop Grid.

The high performance potential of LSDS platforms has also raised a significant interest in the industry. Performance demanding users are also interested by these platforms, considering their cost-performance ratio which is even lower than the one of clusters. Thus, several Desktop Grid platforms are daily used in production in large companies in the domains of pharmacology, petroleum, aerospace, etc.

Desktop Grids share with Grid a common objective: to extend the size and accessibility of a computing infrastructure beyond the limit of a single administration domain. In [68], the authors present the similarities and differences between Grid and Global Computing systems. Two important distinguishing parameters are the user community (professional or not) and the resource ownership (who own the resources and who is using them). From the system architecture perspective, we consider two main differences: the system scale and the lack of control of the participating resources. These two aspects have many consequences, at least on the architecture of system components, the deployment methods, programming models, security (trust) and more generally on the theoretical properties achievable by the system.

Beside Desktop Grids and Grids, large scale parallel computers with tens of thousands (and even hundreds of thousands) of CPU cores are emerging with scalability issues similar to the one of Internet Computing systems: fault tolerance at large scale, large scale data movements, tools and languages. Grand-Large is gradually considering the application of selected research results, in the domain of large scale parallel computers, in particular for the fault tolerance and language topics.

3.1.2. Building a Large Scale Distributed System

This set of studies considers the XtremWeb project as the basis for research, development and experimentation. This LSDS middleware is already operational. This set gathers 4 studies aiming at improving the mechanisms and enlarging the functionalities of LSDS dedicated to computing. The first study considers the architecture of the resource discovery engine which, in principle, is close to an indexing system. The second study concerns the storage and movements of data between the participants of a LSDS. In the third study, we address the issue of scheduling in LSDS in the context of multiple users and applications. Finally the last study seeks to improve the performance and reduce the resource cost of the MPICH-V fault tolerant MPI for desktop grids.

3.1.2.1. The resource discovery engine

A multi-users/multi-applications LSDS for computing would be in principle very close to a P2P file sharing system such as Napster [87], Gnutella [87] and Kazaa [76], except that the shared resource is the CPUs instead of files. The scale and lack of control are common features of the two kinds of systems. Thus, it is likely that solutions sharing fundamental mechanisms will be adopted, such as lower level communication protocols, resource publishing, resource discovery and distributed coordination. As an example, recent P2P projects have proposed distributed indexing systems like CAN [84], CHORD [89], PASTRY [85] and TAPESTRY [94] that could be used for resource discovery in a LSDS dedicated to computing.

The resource discovery engine is composed of a publishing system and a discovery engine, which allow a client of the system to discover the participating nodes offering some desired services. Currently, there is as much resource discovery architectures as LSDS and P2P systems. The architecture of a resource discovery engine is derived from some expected features such as speed of research, speed of reconfiguration, volatility tolerance, anonymity, limited use of the network, matching between the topologies of the underlying network and the virtual overlay network.

This study focuses on the first objective: to build a highly reliable and stable overlay network supporting the higher level services. The overlay network must be robust enough to survive unexpected behaviors (like malicious behaviors) or failures of the underlying network. Unfortunately it is well known that under specific assumptions, a system cannot solve even simples tasks with malicious participants. So, we focus the study on designing overlay algorithms for transient failures. A transient failure accepts any kind of behavior from the system, for a limited time. When failures stop, the system will eventually provide its normal service again.

A traditional way to cope with transient failures are self-stabilizing systems [66]. Existing self-stabilizing algorithms use an underlying network that is not compatible with LSDS. They assume that processors know their list of neighbors, which does not fit the P2P requirements. Our work proposes a new model for designing self-stabilizing algorithms without making this assumption, then we design, prove and evaluate overlay networks self-stabilizing algorithms in this model.

3.1.2.2. Fault Tolerant MPI

MPICH-V is a research effort with theoretical studies, experimental evaluations and pragmatic implementations aiming to provide a MPI implementation based on MPICH [79], featuring multiple fault tolerant protocols.

There is a long history of research in fault tolerance for distributed systems. We can distinguish the automatic/transparent approach from the manual/user controlled approach. The first approach relies either on coordinated checkpointing (global snapshot) or uncoordinated checkpointing associated with message logging. A well known algorithm for the first approach has been proposed by Chandy and Lamport [62]. This algorithm requires restarting all processes even if only one process crashes. So it is believed not to scale well. Several strategies have been proposed for message logging: optimistic [91], pessimistic [48], causal [92]. Several optimizations have been studied for the three strategies. The general context of our study is high performance computing on large platforms. One of the most used programming environments for such platforms is MPI.

Within the MPICH-V project, we have developed and published several original fault tolerant protocols for MPI: MPICH-V1 [55], MPICH-V2 [56], MPICH-Vcausal, MPICH-Vcl [57], MPICH-Pcl. The two first protocols rely on uncoordinated checkpointing associated with either remote pessimistic message logging or sender based pessimistic message logging. We have demonstrated that MPICH-V2 outperforms MPICH-V1. MPICH-Vcl implements a coordinated checkpoint strategy (Chandy-Lamport) removing the need of message logging. MPICH-V2 and Vcl are concurrent protocols for large clusters. We have compared them considering a new parameter for evaluating the merits of fault tolerant protocols: the impact of the fault frequency on the performance. We have demonstrated that the stress of the checkpoint server is the fundamental source of performance differences between the two techniques. MPICH-Vcausal implements a causal message logging protocols, removing the need for waiting acknowledgement in contrary to MPICH-V2. MPICH-Pcl is a blocking implementation of the Vcl protocol. Under the considered experimental conditions, message logging becomes more relevant than coordinated checkpoint when the fault frequency reaches 1 fault every 4 hours, for a cluster of 100 nodes sharing a single checkpoint server, considering a data set of 1 GB on each node and a 100 Mb/s network.

Multiple important events arose from this research topic. A new open source implementation of the MPI-2 standard was born during the evolution of the MPICH-V project, namely OpenMPI. OpenMPI is the result of the alliance of many MPI projects in the USA, and we are working to port our fault tolerance algorithms both into OpenMPI and MPICH.

Grids becoming more popular and accessible than ever, parallel applications developers now consider them as possible targets for computing demanding applications. MPI being the de-facto standard for the programming of parallel applications, many projects of MPI for the Grid appeared these last years. We contribute to this new way of using MPI through a European Project in which we intend to grid-enable OpenMPI and provide new fault-tolerance approaches fitted for the grid.

When introducing Fault-Tolerance in MPI libraries, one of the most neglected component is the runtime environment. Indeed, the traditional approach consists in restarting the whole application and runtime environment

in case of failure. A more efficient approach could be to implement a fault-tolerant runtime environment, capable of coping with failures at its level, thus avoiding the restart of this part of the application. The benefits would be a quicker restart time, and a better control of the application. However, in order to build a fault-tolerant runtime environment for MPI, new topologies, more connected, and more stable, must be integrated in the runtime environment.

For traditional parallel machines of large scale (like large scale clusters), we also continue our investigation of the various fault tolerance protocols, by designing, implementing and evaluating new protocols in the MPICH-V project.

3.2. Volatility and Reliability Processing

In a global computing application, users voluntarily lend the machines, during the period they don't use them. When they want to reuse the machines, it is essential to give them back immediately. We assume that there is no time for saving the state of the computation (for example because the user is shooting down is machine). Because the computer may not be available again, it is necessary to organize checkpoints. When the owner takes control of his machine, one must be able to continue the computation on another computer from a checkpoint as near as possible from the interrupted state.

The problems raised by this way of managing computations are numerous and difficult. They can be put into two categories: synchronization and repartition problems.

- Synchronization problems (example). Assume that the machine that is supposed to continue the computation is fixed and has a recent checkpoint. It would be easy to consider that this local checkpoint is a component of a global checkpoint and to simply rerun the computation. But on one hand the scalability and on the other hand the frequency of disconnections make the use of a global checkpoint totally unrealistic. Then the checkpoints have to be local and the problem of synchronizing the recovery machine with the application is raised.

- Repartition problems (example). As it is also unrealistic to wait for the computer to be available again before rerunning the interrupted application, one has to design a virtual machine organization, where a single virtual machine is implemented as several real ones. With too few real machines for a virtual one, one can produce starvation; with too many, the efficiency is not optimal. The good solution is certainly in a dynamic organization.

These types of problems are not new ([69]). They have been studied deeply and many algorithmic solutions and implementations are available. What is new here and makes these old solutions not usable is scalability. Any solution involving centralization is impossible to use in practice. Previous works validated on former networks can not be reused.

3.2.1. Reliability Processing

We voluntarily presented in a separate section the volatility problem because of its specificity both with respect to type of failures and to frequency of failures. But in a general manner, as any distributed system, a global computing system has to resist to a large set of failures, from crash failures to Byzantine failures, that are related to incorrect software or even malicious actions (unfortunately, this hypothesis has to be considered as shown by DECRYPTHON project or the use of erroneous clients in SETI@HOME project), with in between, transient failures such as loss of message duplication. On the other hand, failures related accidental or malicious memory corruptions have to be considered because they are directly related to the very nature of the Internet. Traditionally, two approaches (masking and non-masking) have been used to deal with reliability problems. A masking solution hides the failures to the user, while a non-masking one may let the user notice that failures occur. Here again, there exists a large literature on the subject (cf. [78], [90], [66] for surveys). Masking techniques, generally based on consensus, are not scalable because they systematically use generalized broadcasting. The self-stabilizing approach (a non-masking solution) is well adapted (specifically its time adaptive version, cf. [75], [74], [52], [53], [70]) for three main reasons:

1. Low overhead when stabilized. Once the system is stabilized, the overhead for maintaining correction is low because it only involves communications between neighbours.

2. Good adaptivity to the reliability level. Except when considering a system that is continuously under attacks, self-stabilization provides very satisfying solutions. The fact that during the stabilization phase, the correctness of the system is not necessarily satisfied is not a problem for many kinds of applications.

3. Lack of global administration of the system. A peer to peer system does not admit a centralized administrator that would be recognized by all components. A human intervention is thus not feasible and the system has to recover by itself from the failures of one or several components, that is precisely the feature of self-stabilizing systems.

We propose:

1. To study the reliability problems arising from a global computing system, and to design self-stabilizing solutions, with a special care for the overhead.

2. For problem that can be solved despite continuously unreliable environment (such as information retrieval in a network), to propose solutions that minimize the overhead in space and time resulting from the failures when they involve few components of the system.

3. For most critical modules, to study the possibility to use consensus based methods.

4. To build an adequate model for dealing with the trade-off between reliability and cost.

3.3. Parallel Programming on Peer-to-Peer Platforms (P5)

Several scientific applications, traditionally computed on classical parallel supercomputers, may now be adapted for geographically distributed heterogeneous resources. Large scale P2P systems are alternative computing facilities to solve grand challenge applications.

Peer-to-Peer computing paradigm for large scale scientific and engineering applications is emerging as a new potential solution for end-user scientists and engineers. We have to experiment and to evaluate such programming to be able to propose the larger possible virtualization of the underlying complexity for the end-user.

3.3.1. *Large Scale Computational Sciences and Engineering*

Parallel and distributed scientific application developments and resource managements in these environments are a new and complex undertaking. In scientific computation, the validity of calculations, the numerical stability, the choices of methods and software are depending of properties of each peer and its software and hardware environments; which are known only at run time and are non-deterministic. The research to obtain acceptable frameworks, methodologies, languages and tools to allow end-users to solve accurately their applications in this context is capital for the future of this programming paradigm.

GRID scientific and engineering computing exists already since more than a decade. Since the last few years, the scale of the problem sizes and the global complexity of the applications increase rapidly. The scientific simulation approach is now general in many scientific domains, in addition to theoretical and experimental aspects, often link to more classic methods. Several applications would be computed on world-spread networks of heterogeneous computers using some web-based Application Server Provider (ASP) dedicated to targeted scientific domains. New very strategic domains, such as Nanotechnologies, Climatology or Life Sciences, are in the forefront of these applications. The development in this very important domain and the leadership in many scientific domains will depend in a close future to the ability to experiment very large scale simulation on adequate systems [73]. The P2P scientific programming is a potential solution, which is based on existing computers and networks. The present scientific applications on such systems are only concerning problems which are mainly data independents: i.e. each peer does not communicate with the others.

P2P programming has to develop parallel programming paradigms which allow more complex dependencies between computing resources. This challenge is an important goal to be able to solve large scientific applications. The results would also be extrapolated toward future petascale heterogeneous hierarchically designed supercomputers.

3.3.2. Experimentations and Evaluations

We have followed two tracks. First, we did experiments on large P2P platforms in order to obtain a realistic evaluation of the performance we can expect. Second, we have set some hypothesis on peers, networks, and scheduling in order to have theoretical evaluations of the potential performance. Then, we have chosen a classical linear algebra method well-adapted to large granularity parallelism and asynchronous scheduling: the block Gauss-Jordan method to invert dense very large matrices. We have also chosen the calculation of one matrix polynomial, which generates computation schemes similar to many linear algebra iterative methods, well-adapted for very large sparse matrices. Thus, we were able to theoretically evaluate the potential throughput with respect to several parameters such as the matrix size and the multicast network speed.

Since the beginning of the evaluations, we experimented with those parallel methods on a few dozen peer XtremWeb P2P Platforms. We continue these experiments on larger platforms in order to compare these results to the theoretical ones. Then, we would be able to extrapolate and obtain potential performance for some scientific applications.

Recently, we also experimented several Krylov based method, such as the Lanczos and GMRES methods on several grids, such as a French-Japanese grid using hundred of PC in France and 4 clusters at the University of Tsukuba. We also experimented on GRID5000 the same methods. We currently use several middleware such as Xtremweb, OmniRPC and Condor. We also begin some experimentations on the Tsubame supercomputer in collaboration with the TITech (Tokyo Institute of Technologies) in order to compare our grid approaches and the High performance one on an hybrid supercomputer.

Experimentations and evaluation for several linear algebra methods for large matrices on P2P systems will always be developed all along the Grand Large project, to be able to confront the different results to the reality of the existing platforms.

As a challenge, we would like, in several months, to efficiently invert a dense matrix of size one million using a several thousand peer platform. We are already inverting very large dense matrices on Grid5000 but more efficient scheduler and a larger number of processors are required to this challenge.

Beyond the experimentations and the evaluations, we propose the basis of a methodology to efficiently program such platforms, which allow us to define languages, tools and interface for the end-user.

3.3.3. Languages, Tools and Interface

The underlying complexity of the Large Scale P2P programming has to be mainly virtualized for the end-user. We have to propose an interface between the end-user and the middleware which may extract the end-user expertise or propose an on-the-shelf general solution. Targeted applications concern very large scientific problems which have to be developed using component technologies and up-to-dated software technologies.

We introduced the YML framework and language which allows to describe dependencies between components. We introduced different classes of components, depending of the level of abstraction, which are associated with divers parts of the framework. A component catalogue is managed by an administrator and/or the end-users. Another catalogue is managed with respect to the experimental platform and the middleware criteria. A front-end part is completely independent of any middleware or testbed, and a back-end part is developed for each targeted middleware/platform couple. A YML scheduler is adapted for each of the targeted systems.

The YML framework and language propose a solution to develop scientific applications to P2P and GRID platform. An end-user can directly develop programs using this framework. Nevertheless, many end-users would prefer avoid programming at the component and dependency graph level. Then, an interface has to be proposed soon, using the YML framework. This interface may be dedicated to a special scientific domain to be able to focus on the end-user vocabulary and P2P programming knowledge. We plan to develop such version based on the YML framework and language. The first targeted scientific domain will be very large linear algebra for dense or sparse matrices.

3.4. Methodology for Large Scale Distributed Systems

Research in the context of LSDS involves understanding large scale phenomena from the theoretical point of view up to the experimental one under real life conditions.

One key aspects of the impact of large scale on LSDS is the emergence of phenomena which are not co-ordinated, intended or expected. These phenomena are the results of the combination of static and dynamic features of each component of LSDS: nodes (hardware, OS, workload, volatility), network (topology, conges-tion, fault), applications (algorithm, parameters, errors), users (behavior, number, friendly/aggressive).

Validating current and next generation of distributed systems targeting large-scale infrastructures is a complex task. Several methodologies are possible. However, experimental evaluations on real testbeds are unavoidable in the life-cycle of a distributed middleware prototype. In particular, performing such real experiments in a rigorous way requires to benchmark developed prototypes at larger and larger scales. Fulfilling this requirement is mandatory in order to fully observe and understand the behaviors of distributed systems. Such evaluations are indeed mandatory to validate (or not!) proposed models of these distributed systems, as well as to elaborate new models. Therefore, to enable an experimentally-driven approach for the design of next generation of large scale distributed systems, developing appropriate evaluation tools is an open challenge.

Fundamental aspects of LSDS as well as the development of middleware platforms are already existing in Grand-Large. Grand-Large aims at gathering several complementary techniques to study the impact of large scale in LSDS: observation tools, simulation, emulation and experimentation on real platforms.

3.4.1. Observation tools

Observation tools are mandatory to understand and extract the main influencing characteristics of a distributed system, especially at large scale. Observation tools produce data helping the design of many key mechanisms in a distributed system: fault tolerance, scheduling, etc. We pursue the objective of developing and deploying a large scale observation tool (XtremLab) capturing the behavior of thousands of nodes participating to popular Desktop Grid projects. The collected data will be stored, analyzed and used as reference in a simulator (SIMBOINC).

3.4.2. Tool for scalability evaluations

Several Grid and P2P systems simulators have been developed by other teams: SimGrid [61], GridSim [59], Briks [46]. All these simulators considers relatively small scale Grids. They have not been designed to scale and simulate 10 K to 100 K nodes. Other simulators have been designed for large multi-agents systems such as Swarm [80] but many of them considers synchronous systems where the system evolution is guided by phases. In the P2P field, ad hoc many simulators have been developed, mainly for routing in DHT. Emulation is another tool for experimenting systems and networks with a higher degree of realism. Compared to simulation, emulation can be used to study systems or networks 1 or 2 orders of magnitude smaller in terms of number of components. However, emulation runs the actual OS/middleware/applications on actual platform. Compared to real testbed, emulation considers conducting the experiments on a fully controlled platform where all static and dynamic parameters can be controlled and managed precisely. Another advantage of emulation over real testbed is the capacity to reproduce experimental conditions. Several implementations/configurations of the system components can be compared fairly by evaluating them under the similar static and dynamic conditions. Grand-Large is leading one of the largest Emulator project in Europe called Grid explorer (French funding). This project has built and used a 1K CPUs cluster as hardware platform and gathers 24 experiments of 80 researchers belonging to 13 different laboratories. Experiments concerned developing the emulator itself and use of the emulator to explore LSDS issues. In term of emulation tool, the main outcome of Grid explorer is the V-DS system, using virtualization techniques to fold a virtual distributed system 50 times larger than the actual execution platform. V-DS aims at discovering, understanding and managing implicit uncoordinated large scale phenomena. Grid Explorer is still in use within the Grid'5000 platform and serves the community of 400 users 7 days a week and 24h a day.

3.4.3. Real life testbeds: extreme realism

The study of actual performance and connectivity mechanisms of Desktop Grids needs some particular testbed where actual middleware and applications can be run under real scale and real life conditions. Grand-Large is

developing DSL-Lab, an experimental platform distributed on 50 sites (actual home of the participants) and using the actual DSL network as the connection between the nodes. Running experiments over DSL-Lab put the piece of software to study under extremely realistic conditions in terms of connectivity (NAT, Firewalls), performance (node and network), performance symmetry (DSL Network is not symmetric), etc.

To investigate real distributed system at large scale (Grids, Desktop Grids, P2P systems), under real life conditions, only a real platform (featuring several thousands of nodes), running the actual distributed system can provide enough details to clearly understand the performance and technical limits of a piece of software. Grand-Large members are strongly involved (as Project Director) in the French Grid5000 project which intents to deploy an experimental Grid testbed for computer scientists. This testbed features about 4000 CPUs gathering the resources of about 9 clusters geographically distributed over France. The clusters will be connected by a high speed network (Renater 10G). Grand-Large is the leading team in Grid5000, chairing the steering committee. As the Principal Investigator of the project, Grand-Large has taken some strong design decisions that nowadays give a real added value of Grid5000 compared to all other existing Grids: reconfiguration and isolation. From these two features, Grid5000 provides the capability to reproduce experimental conditions and thus experimental results, which is the cornerstone of any scientific instrument.

3.5. High Performance Scientific Computing

This research is in the area of high performance scientific computing, and in particular in parallel matrix algorithms. This is a subject of crucial importance for numerical simulations as well as other scientific and industrial applications, in which linear algebra problems arise frequently. The modern numerical simulations coupled with ever growing and more powerful computational platforms have been a major driving force behind a progress in numerous areas as different as fundamental science, technical/technological applications, life sciences.

The main focus of this research is on the design of efficient, portable linear algebra algorithms, such that solving a large set of linear equations or a least squares problem. The characteristics of the matrices commonly encountered in this situations can vary significantly, as are the computational platforms used for the calculations. Nonetheless two common trends are easily discernible. First, the problems to solve are larger and larger, since the numerical simulations are using higher resolution. Second, the architecture of today's supercomputers is getting very complex, and so the developed algorithms need to be adapted to these new achitectures.

3.5.1. *Communication avoiding algorithms for numerical linear algebra*

Since 2007, we work on a novel approach to dense and sparse linear algebra algorithms, which aims at minimizing the communication, in terms of both its volume and a number of transferred messages. This research is motivated by technological trends showing an increasing communication cost. Its main goal is to reformulate and redesign linear algebra algorithms so that they are optimal in an amount of the communication they perform, while retaining the numerical stability. The work here involves both theoretical investigation and practical coding on diverse computational platforms. We refer to the new algorithms as *communication avoiding algorithms* [18] [8]. In our team we focus on communication avoiding algorithms for dense direct methods as well as sparse iterative methods.

The theoretical investigation focuses on identifying lower bounds on communication for different operations in linear algebra, where communication refers to data movement between processors in the parallel case, and to data movement between different levels of memory hierarchy in the sequential case. The lower bounds are used to study the existing algorithms, understand their communication bottlenecks, and design new algorithms that attain them.

This research focuses on the design of linear algebra algorithms that minimize the cost of communication. Communication costs include both latency and bandwidth, whether between processors on a parallel computer or between memory hierarchy levels on a sequential machine. The stability of the new algorithms represents an important part of this work.

3.5.2. *Preconditioning techniques*

Solving a sparse linear system of equations is the most time consuming operation at the heart of many scientific applications, and therefore it has received a lot of attention over the years. While direct methods are robust, they are often prohibitive because of their time and memory requirements. Iterative methods are widely used because of their limited memory requirements, but they need an efficient preconditioner to accelerate their convergence. In this direction of research we focus on preconditioning techniques for solving large sparse systems.

One of the main challenges that we address is the scalability of existing methods as incomplete LU factorizations or Schwarz-based approaches, for which the number of iterations increases significantly with the problem size or with the number of processors. This is often due to the presence of several low frequency modes that hinder the convergence of the iterative method. To address this problem, we study direction preserving solvers in the context of multilevel filtering LU decompositions. A judicious choice for the directions to be preserved through filtering allows us to alleviate the effect of low frequency modes on the convergence. While preconditioners and their scalability are studied by many other groups, our approach of direction preserving and filtering is studied in only very few other groups in the world (as Lawrence Livermore National Laboratory, Frankfurt University, Pennsylvania State University).

3.5.3. *Fast linear algebra solvers based on randomization*

Linear algebra calculations can be enhanced by statistical techniques in the case of a square linear system $Ax = b$ where A is a general or symmetric indefinite matrix [16]& [25]. Thanks to a random transformation of A, it is possible to avoid pivoting and then to reduce the amount of communication. Numerical experiments show that this randomization can be performed at a very affordable computational price while providing us with a satisfying accuracy when compared to partial pivoting. This random transformation called Partial Random Butterfly Transformation (PRBT) is optimized in terms of data storage and flops count. A PRBT solver for LU factorization (and for LDL^T factorization on multicore) has been developed. This solver takes advantage of the latest generation of hybrid multicore/GPU machines and gives better Gflop/s performance than existing factorization routines.

3.5.4. *Sensitivity analysis of linear algebra problems*

We derive closed formulas for the condition number of a linear function of the total least squares solution [2]. Given an over determined linear systems $Ax = b$, we show that this condition number can be computed using the singular values and the right singular vectors of $[A, b]$ and A. We also provide an upper bound that requires the computation of the largest and the smallest singular value of $[A, b]$ and the smallest singular value of A. In numerical experiments, we compare these values with condition estimates from the literature.

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

3. Scientific Foundations

3.1. Analytical information theory

Participants: Cédric Adjih, Pascale Minet, Paul Mühlethaler.

channel capacity, compression, predictors

> **Information theory** Branch of mathematics dedicated to the quantification of the performance of a medium to carry information. Initiated by Shannon in 1948.

Abstract. Information theory and analytical methods play a central role in the networking technology. It identifies the key parameter that must be quantified in order to characterize the performance of a network.

The analytical information theory is part of the foundations of the Hipercom project. This is a tool box that has been collected and adapted from the areas of the analysis of algorithms and the information theory. It provides powerful tool for the analysis of telecommunication algorithms. The analysis of the behavior of such algorithms in their asymptotic range are fundamental in order to identify their critical parts. It helps to design and properly scale the protocols. Application of analytical information theory ranges from channel capacity computations, compression algorithm performance evaluation, predictor designs.

3.2. Methodology of telecommunication algorithm evaluation

Participants: Cédric Adjih, Ichrak Amdouni, Emmanuel Baccelli, Salman Malik, Yacine Mezali, Pascale Minet, Paul Mühlethaler, Saoucene Mahfoudh Ridene, Ridha Soua, Erwan Livolant, Ines Khoufi.

deterministic performance, probabilistic performance

> **Power laws** probability distributions that decays has inverse power of the variable for large values of the variable. Power laws are frequent in economic and statistical analysis (see Pareto law). Simple models such as Poisson processes and finite state Markov processes don't generate distributions with power laws.

We develop our performance evaluation tools towards deterministic performance and probabilistic performance. Our tools range from mathematical analysis to simulation and real life experiment of telecommunication algorithms.

One cannot design good algorithms without good evaluation models. Hipercom project team has an historically strong experience in performance evaluation of telecommunication systems, notably when they have multiple access media. We consider two main methodologies:

- Deterministic performance analysis,
- Probabilistic performance analysis

In the deterministic analysis, the evaluation consists to identify and quantify the worst case scenario for an algorithm in a given context. For example to evaluate an end-to-end delay. Mathematically it consists into handling a (max,+) algebra. Since such algebra is not commutative, the complexity of the evaluation of an end-to-end delay frequently grows exponentially with the number of constraints. Therefore the main issue in the deterministic evaluation of performance is to find bounds easier to compute in order to have practical results in realistic situations.

In the probabilistic analysis of performance, one evaluate the behavior of an algorithm under a set of parameters that follows a stochastic model. For example traffic may be randomly generated, nodes may move randomly on a map. The pionneer works in this area come from Knuth (1973) who has systemized this branch. In the domain of telecommunication, the domain has started a significant rise with the appearance of the problematic of collision resolution in a multiple access medium. With the rise of wireless communication, new interesting problems have been investigated.

The analysis of algorithm can rely on analytical methodology which provides the better insight but is practical in very simplistic models. Simulation tools can be used to refine results in more complicated models. At the end of the line, we proceed with real life experiments. To simplify, experiments check the algorithms with 10 nodes in maximum, simulations with 100 nodes maximum, analytical tools with more 1,000 nodes, so that the full range of applicability of the algorithms is investigated.

3.3. Traffic and network architecture modeling

Participants: Cédric Adjih, Aline Carneiro Viana, Emmanuel Baccelli.

traffic source models, network topologies, mobility models, dynamic nodes

Abstract. Network models are important. We consider four model problems: topology, mobility, dynamics and traffic models.

One needs good and realistic models of communication scenarios in order to provide pertinent performance evaluation of protocols. The models must assess the following key points:

- The architecture and topology: the way the nodes are structured within the network
- The mobility: the way the nodes move
- The dynamics: the way the nodes change status
- The traffic: the way the nodes communicate

For the architecture there are several scales. At the internet scale it is important to identify the patterns which dictate the node arrangement. For example the internet topology involves many power law distribution in node degree, link capacities, round trip delays. These parameters have a strong impact in the performance of the global network. At a smaller scale there is also the question how the nodes are connected in a wireless network. There is a significant difference between indoor and outdoor networks. The two kinds of networks differ on wave propagation. In indoor networks, the obstacles such as walls, furniture, etc, are the main source of signal attenuations. In outdoor networks the main source of signal attenuation is the distance to the emitter. This lead to very different models which vary between the random graph model for indoor networks to the unit graph model for outdoor networks.

The mobility model is very important for wireless network. The way nodes move may impact the performance of the network. For example it determines when the network splits in distinct connected components or when these components merge. With random graph models, the mobility model can be limited to the definition of a link status holding time. With unit disk model the mobility model will be defined according to random speed and direction during random times or random distances. There are some minor complications on the border of the map.

The node dynamic addresses the elements that change inside the node. For example its autonomy, its bandwidth requirement, the status of server, client, etc. Pair to pair networks involve a large class of users who frequently change status. In a mobile ad hoc network, nodes may change status just by entering a coverage area, or because some other nodes leaves the coverage area.

The traffic model is very most important. There are plenty literature about trafic models which arose when Poisson models was shown not to be accurate for real traffics, on web or on local area networks. Natural traffic shows long range dependences that don't exist in Poisson traffic. There are still strong issues about the origin of this long range dependences which are debated, however they have a great impact on network performance since congestions are more frequent. The origin are either from the distribution of file sizes exchanged over the net, or from the protocols used to exchange them. One way to model the various size is to consider on/off sources. Every time a node is on it transfers a file of various size. The TCP protocol has also an impact since it keeps a memory on the network traffic. One way to describe it is to use an on/off model (a source sending packets in transmission windows) and to look at the superposition of these on/off sources.

3.4. Algorithm design, evaluation and implementation

Participants: Cédric Adjih, Aline Carneiro Viana, Emmanuel Baccelli, Saoucene Mahfoudh Ridene, Pascale Minet, Paul Mühlethaler, Ridha Soua, Erwan Livolant, Ines Khoufi.

Access protocols, routing, scheduling, QoS

Abstract. Algorithms are conceived with focal point on performance. The algorithms we specify in detail range between medium access control to admission control and quality of service management.

The conception of algorithms is an important focus of the project team. We specify algorithms in the perspective of achieving the best performance for communication. We also strive to embed those algorithms in protocols that involve the most legacy from existing technologies (Operating systems, internet, Wifi). Our aim with this respect is to allow code implementations for real life experiment or imbedded simulation with existing network simulators. The algorithm specified by the project ranges from multiple access schemes, wireless ad hoc routing, mobile multicast management, Quality of service and admission controls. In any of these cases the design emphasize the notions of performance, robustness and flexibility. For example, a flooding technique in mobile ad hoc network should be performing such to save bandwidth but should not stick too much close to optimal in order to be more reactive to frequent topology changes. Some telecommunication problems have NP hard optimal solution, and an implementable algorithm should be portable on very low power processing unit (e.g. sensors). Compromise are found are quantified with respect to the optimal solution.

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

3. Scientific Foundations

3.1. Scientific Foundations

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

The use of graphic representation to understand abstract data is a goal Visual Analytics shares with Tukey's Exploratory Data Analysis (EDA) [62], graphic designers such as Bertin [51] and Tufte [61], and HCI researchers in the field of Information Visualization [50].

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

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

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

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

Several theories of perception are related to information visualization such as the "Gestalt" principles, Gibson's theory of visual perception [56] and Triesman's "preattentive processing" theory [60]. We use them extensively but they only have a limited accuracy for predicting the effectiveness of novel visual representations in interactive settings.

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

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

In 2002, we showed that the practical limit of InfoVis was on the order of 1 million items displayed on a screen [54]. Although screen technologies have improved rapidly since then, eventually we will be limited by the physiology of our vision system: about 20 millions receptor cells (rods and cones) on the retina. Another problem will be the limits of human visual attention, as suggested by our 2006 study on change blindness in large and multiple displays [52]. Therefore, visualization alone cannot let us understand very large data sets. Other techniques such as aggregation or sampling must be used to reduce the visual complexity of the data to the scale of human perception.

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

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

Designing novel interaction and visualization techniques to explore huge data sets is an important goal and requires solving hard problems, but how can we assess whether or not our techniques and systems provide real improvements? Without this answer, we cannot know if we are heading in the right direction. This is why we have been actively involved in the design of evaluation methods for information visualization [8] [59], [57], [58], [55]. For more complex systems, other methods are required. For these we want to focus on longitudinal evaluation methods while still trying to improve controlled experiments.

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

3. Scientific Foundations

3.1. Scientific Foundations

Dahu has strong connections with the Leo project-team in Saclay.

Dahu aims at developing mechanisms for high-level specifications of systems built around DBMS, that are easy to understand while also facilitating verification of critical properties. This requires developing tools that are suitable for reasoning about systems that manipulate data. Some tools for specifying and reasoning about data have already been studied independently by the database community and by the verification community, with various motivations. However, this work is still in its infancy and needs to be further developed and unified.

Most current proposals for reasoning about DBMS over XML documents are based on tree automata, taking advantage of the tree structure of XML documents. For this reason, the Dahu team is studying a variety of tree automata. This ranges from restrictions of "classical" tree automata in order to understand their expressive power, to extensions of tree automata in order to understand how to incorporate the manipulation of data.

Moreover, Dahu is also interested in logical frameworks that explicitly refer to data. Such logical frameworks can be used as high level declarative languages for specifying integrity constraints, format change during data exchange, web service functionalities and so on. Moreover, the same logical frameworks can be used to express the critical properties we wish to verify.

In order to achieve its goals, Dahu brings together world-class expertise in both databases and verification.

<p style="text-align:center;color:red;font-weight:bold;">IN-SITU Project-Team</p>

3. Scientific Foundations

3.1. Multi-disciplinary Research

INSITU uses a multi-disciplinary research approach, including computer scientists, psychologists and designers. Working together requires an understanding of each other's methods. Much of computer science relies on formal theory, which, like mathematics, is evaluated with respect to its internal consistency. The social sciences are based more on descriptive theory, attempting to explain observed behaviour, without necessarily being able to predict it. The natural sciences seek predictive theory, using quantitative laws and models to not only explain, but also to anticipate and control naturally occurring phenomena. Finally, design is based on a corpus of accumulated knowledge, which is captured in design practice rather than scientific facts but is nevertheless very effective.

Combining these approaches is a major challenge. We are exploring an integrative approach that we call *generative theory*, which builds upon existing knowledge in order to create new categories of artefacts and explore their characteristics Our goal is to produce prototypes, research methods and software tools that facilitate the design, development and evaluation of interactive systems [40].

OAK Team

3. Scientific Foundations

3.1. Efficient XML and RDF data management

The development of Web technologies has led to a strong increase in the number and complexity of the applications which represent their data in Web formats, among which XML (for structured documents) and RDF (for Semantic Web data) are the most prominent. Oak has carried on research on algorithms and systems for efficiently processing expressive queries on such Web data formats. We have considered the efficient management of XML and RDF data, both for query evaluation and for efficiently applying updates, possibly in concurrence with queries. We have also started investigating multidimensional data analysis within RDF data warehouses.

For applications that integrate such Web data from various sources, we developed efficient and effective techniques to automatically recognize multiple representations of the same real-world object. That is, we devised main-memory resident algorithms that apply on hierarchical data [8] as well as algorithms that manipulate graph data leveraging off-the-shelf database management systems and parallelization to address both efficiency and scalability beyond main-memory [7].

3.2. Cloud-based Data Management

We have recently started to work on the efficient management of complex Web data, in particular structured XML documents and Semantic Web data under the form of RDF, in a cloud-based data management platform. We have investigated architectures and algorithms for storing Web data in elastic cloud-based stores and building an index for such data within the efficient key-value stores provided by off-the-shelf cloud data management platform. We have devised and prototyped such platforms for both XML and RDF data, and started experimenting with them in the Amazon Web Services (AWS) platform [13], [12], [10].

3.3. Data Transformation Management

With the increasing complexity of data processing queries, for instance in applications such as relational data analysis or integration of Web data (e.g., XML or RDF) comes the need to better manage complex data transformations. This includes systematically verifying, maintaining, and testing the transformations an application relies on. In this context, Oak has focused on verifying the semantic correctness of a declarative program that specifies a data transformation query, e.g., an SQL query. To this end, we have investigated how to leverage data provenance (the information of the origin of data and the query operators) for query debugging. More specifically, we developed and implemented algorithms to explain unexpected results produced by a query (why-provenance) as well as expected results that are however missing from the query result (why-not provenance). Results have been presented in form of a software prototype [15].