RESEARCH CENTER
Saclay - Île-de-France

FIELD

Activity Report 2012

Section Application Domains

Edition: 2013-04-24

AMIB Project-Team (section vide)

AVIZ Project-Team

4. Application Domains

4.1. Application Domains

AVIZ develops active collaboration with users from various application domains, making sure it can support their specific needs. By studying similar problems in different domains, we can begin to generalize our results and have confidence that our solutions will work for a variety of applications.

Our current application domains include:

- *Genealogy*, in cooperation with North Carolina State University;
- *Biological research*, in cooperation with Institut Pasteur;
- *Digital Libraries*, in cooperation with the French National Archives and the Wikipedia community;
- *Open Data*, in cooperation with Google Open Data and Data Publica;
- *Agrifood Process Modeling*, in cooperation with the DREAM project (see section 8.2.1.1);

BYMOORE Exploratory Action (section vide)

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Miguel Andrés, Nicolás Bordenabe, Konstantinos Chatzikokolakis, Jérémy Dubreil, Catuscia Palamidessi.

The aim of our research is the specification and verification of protocols used in mobile distributed systems, in particular security protocols. We are especially interested in protocols for *information hiding*.

Information hiding is a generic term which we use here to refer to the problem of preventing the disclosure of information which is supposed to be secret or confidential. The most prominent research areas which are concerned with this problem are those of *secure information flow* and of *privacy*.

Secure information flow refers to the problem of avoiding the so-called *propagation* of secret data due to their processing. It was initially considered as related to software, and the research focussed on type systems and other kind of static analysis to prevent dangerous operations, Nowadays the setting is more general, and a large part of the research effort is directed towards the investigation of probabilistic scenarios and treaths.

Privacy denotes the issue of preventing certain information to become publicly known. It may refer to the protection of *private data* (credit card number, personal info etc.), of the agent's identity (*anonymity*), of the link between information and user (*unlinkability*), of its activities (*unobservability*), and of its *mobility* (*untraceability)*.

The common denominator of this class of problems is that an adversary can try to infer the private information (*secrets*) from the information that he can access (*observables*). The solution is then to obfuscate the link between secrets and observables as much as possible, and often the use randomization, i.e. the introduction of *noise*, can help to achieve this purpose. The system can then be seen as a *noisy channel*, in the information-theoretic sense, between the secrets and the observables.

We intend to explore the rich set of concepts and techniques in the fields of information theory and hypothesis testing to establish the foundations of quantitive information flow and of privacy, and to develop heuristics and methods to improve mechanisms for the protection of secret information. Our approach will be based on the specification of protocols in the probabilistic asynchronous π-calculus, and the application of model-checking to compute the matrices associated to the corresponding channels.

COMMANDS Project-Team

4. Application Domains

4.1. Introduction

Commands is a team with a strong commitment in tackling real-life applications in addition to theoretical challenges. This shows in our long history of contracts with industrial partners. In the recent years, we have mainly contributed to the following fields of application.

4.2. Aerospace applications

In the framework of a long-term partnership with the Cnes, and more recently Astrium, we have studied trajectory optimization for space launcher problems. This kind of problems typically involves hard constraints (thermal flux, mechanical efforts) and inexact models (atmosphere, aerodynamic forces). The two main achievments were to study when singular arcs may occur, and to show the effectiveness of a HJB approach on a reduced model. Singular arcs are flight phases with a non-maximal thrust, induced by a tradeoff between speed and atmospheric drag; they cause difficulties of both theoretical and practical nature. The latter point is the first step in the process of applying global methods to this class of difficult problems.

4.3. Trading applications

In a partnership with Total, we have studied problems dealing with the trading of Liquefied Natural Gas. We have computed maximizing revenue policies, by combining the Stochastic Dual Dynamic Programmin approach (SDDP) with a quantization method for the noise that enters in prices. We have also given partial results for the case of integer decision.

4.4. Energy applications

With Renault, we have studied problems of energy management for hybrid vehicles. Hybrid vehicles include an auxiliary thermal (gas) engine that is used as a range extender for the main electric propulsion. We are interested in determining the optimal policies for energy management, taking into account some stochatic incertainties, as well as execution delay and decision lags.

DAHU Project-Team

4. Application Domains

4.1. Application Domains

Databases are pervasive across many application fields. Indeed, most human activities today require some form of data management. In particular, all applications involving the processing of large amounts of data require the use of a database. Increasingly complex Web applications and services also rely on DBMS, and their correctness and robustness is crucial.

We believe that the automated solutions that Dahu aims to develop for verifying such systems will be useful in this context.

DEFI Project-Team

4. Application Domains

4.1. Radar and GPR applications

Conventional radar imaging techniques (ISAR, GPR, etc.) use backscattering data to image targets. The commonly used inversion algorithms are mainly based on the use of weak scattering approximations such as the Born or Kirchhoff approximation leading to very simple linear models, but at the expense of ignoring multiple scattering and polarization effects. The success of such an approach is evident in the wide use of synthetic aperture radar techniques.

However, the use of backscattering data makes 3-D imaging a very challenging problem (it is not even well understood theoretically) and as pointed out by Brett Borden in the context of airborne radar: “In recent years it has become quite apparent that the problems associated with radar target identification efforts will not vanish with the development of more sensitive radar receivers or increased signal-to-noise levels. In addition it has (slowly) been realized that greater amounts of data - or even additional “kinds” of radar data, such as added polarization or greatly extended bandwidth - will all suffer from the same basic limitations affiliated with incorrect model assumptions. Moreover, in the face of these problems it is important to ask how (and if) the complications associated with radar based automatic target recognition can be surmounted.” This comment also applies to the more complex GPR problem.

Our research themes will incorporate the development, analysis and testing of several novel methods, such as sampling methods, level set methods or topological gradient methods, for ground penetrating radar application (imaging of urban infrastructures, landmines detection, underground waste deposits monitoring, ...) using multistatic data.

4.2. Biomedical imaging

Among emerging medical imaging techniques we are particularly interested in those using low to moderate frequency regimes. These include Microwave Tomography, Electrical Impedance Tomography and also the closely related Optical Tomography technique. They all have the advantage of being potentially safe and relatively cheap modalities and can also be used in complementarity with well established techniques such as X-ray computed tomography or Magnetic Resonance Imaging.

With these modalities tissues are differentiated and, consequentially can be imaged, based on differences in dielectric properties (some recent studies have proved that dielectric properties of biological tissues can be a strong indicator of the tissues functional and pathological conditions, for instance, tissue blood content, ischemia, infarction, hypoxia, malignancies, edema and others). The main challenge for these functionalities is to built a 3-D imaging algorithm capable of treating multi-static measurements to provide real-time images with highest (reasonably) expected resolutions and in a sufficiently robust way.

Another important biomedical application is brain imaging. We are for instance interested in the use of EEG and MEG techniques as complementary tools to MRI. They are applied for instance to localize epileptic centers or active zones (functional imaging). Here the problem is different and consists into performing passive imaging: the epileptic centers act as electrical sources and imaging is performed from measurements of induced currents. Incorporating the structure of the skull is primordial in improving the resolution of the imaging procedure. Doing this in a reasonably quick manner is still an active research area, and the use of asymptotic models would offer a promising solution to fix this issue.

4.3. Non destructive testing and parameter identification

One challenging problem in this vast area is the identification and imaging of defaults in anisotropic media. For instance this problem is of great importance in aeronautic constructions due to the growing use of composite materials. It also arises in applications linked with the evaluation of wood quality, like locating knots in timber in order to optimize timber-cutting in sawmills, or evaluating wood integrity before cutting trees. The anisotropy of the propagative media renders the analysis of diffracted waves more complex since one cannot only relies on the use of backscattered waves. Another difficulty comes from the fact that the micro-structure of the media is generally not well known a priori.

Our concern will be focused on the determination of qualitative information on the size of defaults and their physical properties rather than a complete imaging which for anisotropic media is in general impossible. For instance, in the case of homogeneous background, one can link the size of the inclusion and the index of refraction to the first eigenvalue of so-called interior transmission problem. These eigenvalues can be determined form the measured data and a rough localization of the default. Our goal is to extend this kind of idea to the cases where both the propagative media and the inclusion are anisotropic. The generalization to the case of cracks or screens has also to be investigated.

In the context of nuclear waste management many studies are conducted on the possibility of storing waste in a deep geological clay layer. To assess the reliability of such a storage without leakage it is necessary to have a precise knowledge of the porous media parameters (porosity, tortuosity, permeability, etc.). The large range of space and time scales involved in this process requires a high degree of precision as well as tight bounds on the uncertainties. Many physical experiments are conducted *in situ* which are designed for providing data for parameters identification. For example, the determination of the damaged zone (caused by excavation) around the repository area is of paramount importance since microcracks yield drastic changes in the permeability. Level set methods are a tool of choice for characterizing this damaged zone.

4.4. Diffusion MRI

In biological tissues, water is abundant and magnetic resonance imaging (MRI) exploits the magnetic property of the nucleus of the water proton. The imaging contrast (the variations in the grayscale in an image) in standard MRI can be from either proton density, T1 (spin-lattice) relaxation, or T2 (spin-spin) relaxation and the contrast in the image gives some information on the physiological properties of the biological tissue at different physical locations of the sample. The resolution of MRI is on the order of millimeters: the greyscale value shown in the imaging pixel represents the volume-averaged value taken over all the physical locations contained that pixel.

In diffusion MRI, the image contrast comes from a measure of the average distance the water molecules have moved (diffused) during a certain amount of time. The Pulsed Gradient Spin Echo (PGSE) sequence is a commonly used sequence of applied magnetic fields to encode the diffusion of water protons. The term 'pulsed' means that the magnetic fields are short in duration, an the term gradient means that the magnetic fields vary linearly in space along a particular direction. First, the water protons in tissue are labelled with nuclear spin at a precession frequency that varies as a function of the physical positions of the water molecules via the application of a pulsed (short in duration, lasting on the order of ten milliseconds) magnetic field. Because the precessing frequencies of the water molecules vary, the signal, which measures the aggregate phase of the water molecules, will be reduced due to phase cancellations. Some time (usually tens of milliseconds) after the first pulsed magnetic field, another pulsed magnetic field is applied to reverse the spins of the water molecules. The time between the applications of two pulsed magnetic fields is called the 'diffusion time'. If the water molecules have not moved during the diffusion time, the phase dispersion will be reversed, hence the signal loss will also be reversed, the signal is called refocused. However, if the molecules have moved during the diffusion time, the refocusing will be incomplete and the signal detected by the MRI scanner if weaker than if the water molecules have not moved. This lack of complete refocusing is called the signal attenuation and is the basis of the image contrast in DMRI. the pixels showning more signal attenuation is associated with further water displacement during the diffusion time, which may be linked to physiological factors, such as higher cell membrane permeability, larger cell sizes, higher extra-cellular volume fraction.

We model the nuclear magnetization of water protons in a sample due to diffusion-encoding magnetic fields by a multiple compartment Bloch-Torrey partial differential equation, which is a diffusive-type time-dependent PDE. The DMRI signal is the integral of the solution of the Bloch-Torrey PDE. In a homogeneous medium, the intrinsic diffusion coeffcient D will appear as the slope of the semi-log plot of the signal (in approporiate units). However, because during typical scanning times, $50 - 100ms$, water molecules have had time to travel a diffusion distance which is long compared to the average size of the cells, the slope of the semi-log plot of the signal is in fact a measure of an 'effective' diffusion coefficient. In DMRI applications, this measured quantity is called the 'apparent diffusion coefficient' (ADC) and provides the most commonly used form the image contrast for DMRI. This ADC is closely related to the effective diffusion coefficient obtainable from mathematical homogenization theory.

DISCO Project-Team

4. Application Domains

4.1. Control of engineering systems

The team considers control problems in the aeronautic area and studies delay effects in automatic visual tracking on mobile carriers.

4.2. Analysis and Control of life sciences systems

The team is also involved in life sciences applications. The two main lines are the analysis of bioreactors models and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML).

4.3. Energy Management

The team is interested in Energy management and considers optimization and control problems in energy networks.

GALEN Team

4. Application Domains

4.1. Clinical Projects

- **MR & Muscular Diseases**: The use of MR and Diffusion Tensor Imaging are investigated in collaboration with the Henri Mondor University Hospital and Institut of Myology towards automatic quantification of muscular mass loss and non-invasive biopsy. The aim is to provide tools that could be used to automatically analyze MR imaging and extract useful clinical measurements (Insitut of Myology), and assess the potential impact of diffusion tensor imaging towards automatic quantification either of muscular diseases progression.
- **Image-driven Radiotherapy Treatment & Surgery Guidance** : The use of CT and MR imaging for cancer guidance treatment in collaboration with the Oscar Lambert Center. The aim is to provide tools for automatic dose estimation as well as off-line and on-line positioning guidance through deformable fusion between imaging data corresponding to perioding patient treatment. The same concept will be explored in collaboration with the Saint-Antoine University Hospital towards image-driven surgery guidance through 2D to 3D registration between interventional and pre-operative annotated data.
- **MR Brain Imaging towards Low-Gliomas Tumor Brain Understanding**: The use of contrast enhanced imaging is investigated in collaboration with the Montpellier University Hospital towards better understanding of low-gliomas positioning, automatic tumor segmentation/identification and longitudinal (tumor) growth modeling.
- **CT/MR Perfusion Imaging**: The use of perfusion imaging is investigated in collaboration with the Georges Pompidou European Hospital towards compartmental analysis and measuring tissue perfusion and capillary permeability in liver tumors.

GECO Team

4. Application Domains

4.1. Quantum control

The issue of designing efficient transfers between different atomic or molecular levels is crucial in atomic and molecular physics, in particular because of its importance in those fields such as photochemistry (control by laser pulses of chemical reactions), nuclear magnetic resonance (NMR, control by a magnetic field of spin dynamics) and, on a more distant time horizon, the strategic domain of quantum computing. This last application explicitly relies on the design of quantum gates, each of them being, in essence, an open loop control law devoted to a prescribed simultaneous control action. NMR is one of the most promising techniques for the implementation of a quantum computer.

Physically, the control action is realized by exciting the quantum system by means of one or several external fields, being them magnetic or electric fields. The resulting control problem has attracted increasing attention, especially among quantum physicists and chemists (see, for instance, [88], [93]). The rapid evolution of the domain is driven by a multitude of experiments getting more and more precise and complex (see the recent review [48]). Control strategies have been proposed and implemented, both on numerical simulations and on physical systems, but there is still a large gap to fill before getting a complete picture of the control properties of quantum systems. Control techniques should necessarily be innovative, in order to take into account the physical peculiarities of the model and the specific experimental constraints.

The area where the picture got clearer is given by finite dimensional linear closed models.

- **Finite dimensional** refers to the dimension of the space of wave functions, and, accordingly, to the finite number of energy levels.
- **Linear** means that the evolution of the system for a fixed (constant in time) value of the control is determined by a linear vector field.
- **Closed** refers to the fact that the systems are assumed to be totally disconnected from the environment, resulting in the conservation of the norm of the wave function.

The resulting model is well suited for describing spin systems and also arises naturally when infinite dimensional quantum systems of the type discussed below are replaced by their finite dimensional Galerkin approximations. Without seeking exhaustiveness, let us mention some of the issues that have been tackled for finite dimensional linear closed quantum systems:

- controllability [29],
- bounds on the controllability time [25],
- STIRAP processes [98],
- simultaneous control [71],
- optimal control ([67], [38], [50]),
- numerical simulations [77].

Several of these results use suitable transformations or approximations (for instance the so-called rotating wave) to reformulate the finite-dimensional Schrödinger equation as a sub-Riemannian system. Open systems have also been the object of an intensive research activity (see, for instance, [30], [68], [89], [44]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [34], [45], [63], [35]). The controllability issue is less understood than in the finite dimensional setting, but several advances should be mentioned. First of all, it is known that one cannot expect exact controllability on the whole Hilbert sphere [97]. Moreover, it has been shown that a relevant model, the quantum oscillator, is not even approximately controllable [90], [80]. These negative results have been more recently completed by positive ones. In [36], [37] Beauchard and Coron obtained the first positive controllability result for a quantum particle in a 1D potential well. The result is highly nontrivial and is based on Coron's return method (see [52]). Exact controllability is proven to hold among regular enough wave functions. In particular, exact controllability among eigenfunctions of the uncontrolled Schrödinger operator can be achieved. Other important approximate controllability results have then been proved using Lyapunov methods [79], [84], [64]. While [79] studies a controlled Schrödinger equation in $\mathbb{R}$ for which the uncontrolled Schrödinger operator has mixed spectrum, [84], [64] deal mainly with general discrete-spectrum Schrödinger operators.

In all the positive results recalled in the previous paragraph, the quantum system is steered by a single external field. Different techniques can be applied in the case of two or more external fields, leading to additional controllability results [55], [41].

The picture is even less clear for nonlinear models, such as Gross–Pitaevski and Hartree–Fock equations. The obstructions to exact controllability, similar to the ones mentioned in the linear case, have been discussed in [61]. Optimal control approaches have also been considered [33], [46]. A comprehensive controllability analysis of such models is probably a long way away.

4.2. Neurophysiology

At the interface between neurosciences, mathematics, automatics and humanoid robotics, an entire new approach to neurophysiology is emerging. It arouses a strong interest in the four communities and its development requires a joint effort and the sharing of complementary tools.

A family of extremely interesting problems concerns the understanding of the mechanisms supervising some sensorial reactions or biomechanics actions such as image reconstruction by the primary visual cortex, eyes movement and body motion.

In order to study these phenomena, a promising approach consists in identifying the motion planning problems undertaken by the brain, through the analysis of the strategies that it applies when challenged by external inputs. The role of control is that of a language allowing to read and model neurological phenomena. The control algorithms would shed new light on the brain's geometric perception (the so-called neurogeometry [86]) and on the functional organization of the motor pathways.

- A challenging problem is that of the understanding of the mechanisms which are responsible for the process of image reconstruction in the primary visual cortex V1.

  The visual cortex areas composing V1 are notable for their complex spatial organization and their functional diversity. Understanding and describing their architecture requires sophisticated modeling tools. At the same time, the structure of the natural and artificial images used in visual psychophysics can be fully disclosed only using rather deep geometric concepts. The word "geometry" refers here to the internal geometry of the functional architecture of visual cortex areas (not to the geometry of the Euclidean external space). Differential geometry and analysis both play a fundamental role in the description of the structural characteristics of visual perception.

  A model of human perception based on a simplified description of the visual cortex V1, involving geometric objects typical of control theory and sub-Riemannian geometry, has been first proposed by Petitot ([87]) and then modified by Citti and Sarti ([51]). The model is based on experimental observations, and in particular on the fundamental work by Hubel and Wiesel [60] who received the Nobel prize in 1981.

In this model, neurons of V1 are grouped into orientation columns, each of them being sensitive to visual stimuli arriving at a given point of the retina and oriented along a given direction. The retina is modeled by the real plane, while the directions at a given point are modeled by the projective line. The fiber bundle having as base the real plane and as fiber the projective line is called the *bundle of directions of the plane*.

From the neurological point of view, orientation columns are in turn grouped into hypercolumns, each of them sensitive to stimuli arriving at a given point, oriented along any direction. In the same hypercolumn, relative to a point of the plane, we also find neurons that are sensitive to other stimuli properties, such as colors. Therefore, in this model the visual cortex treats an image not as a planar object, but as a set of points in the bundle of directions of the plane. The reconstruction is then realized by minimizing the energy necessary to activate orientation columns among those which are not activated directly by the image. This gives rise to a sub-Riemannian problem on the bundle of directions of the plane.

- Another class of challenging problems concern the functional organization of the motor pathways.

  The interest in establishing a model of the motor pathways, at the same time mathematically rigorous and biologically plausible, comes from the possible spillovers in neurophysiology. It could help to design better control strategies for robots and artificial limbs, rendering them capable to move more progressively and smoothly and also to react to exterior perturbations in a flexible way. An underlying relevant societal goal (clearly beyond our domain of expertise) is to clarify the mechanisms of certain debilitating troubles such as cerebellar disease, chorea and Parkinson's disease.

  A key issue in order to establish a model of the motor pathways is to determine the criteria underlying the brain's choices. For instance, for the problem of human locomotion (see [32]), identifying such criteria would be crucial to understand the neural pathways implicated in the generation of locomotion trajectories.

  A nowadays widely accepted paradigm is that, among all possible movements, the accomplished ones satisfy suitable optimality criteria (see [96] for a review). One is then led to study an inverse optimal control problem: starting from a database of experimentally recorded movements, identify a cost function such that the corresponding optimal solutions are compatible with the observed behaviors.

  Different methods have been taken into account in the literature to tackle this kind of problems, for instance in the linear quadratic case [66] or for Markov processes [85]. However all these methods have been conceived for very specific systems and they are not suitable in the general case. Two approaches are possible to overcome this difficulty. The direct approach consists in choosing a cost function among a class of functions naturally adapted to the dynamics (such as energy functions) and to compare the solutions of the corresponding optimal control problem to the experimental data. In particular one needs to compute, numerically or analytically, the optimal trajectories and to choose suitable criteria (quantitative and qualitative) for the comparison with observed trajectories. The inverse approach consists in deriving the cost function from the qualitative analysis of the data.

4.3. Switched systems

Switched systems form a subclass of hybrid systems, which themselves constitute a key growth area in automation and communication technologies with a broad range of applications. Existing and emerging areas include automotive and transportation industry, energy management and factory automation. The notion of hybrid systems provides a framework adapted to the description of the heterogeneous aspects related to the interaction of continuous dynamics (physical system) and discrete/logical components.

The characterizing feature of switched systems is the collective aspect of the dynamics. A typical question is that of stability, in which one wants to determine whether a dynamical system whose evolution is influenced by a time-dependent signal is uniformly stable with respect to all signals in a fixed class ([73]).

The theory of finite-dimensional hybrid and switched systems has been the subject of intensive research in the last decade and a large number of diverse and challenging problems such as stabilizability, observability, optimal control and synchronization have been investigated (see for instance [94], [74]).

The question of stability, in particular, because of its relevance for applications, has spurred a rich literature. Important contributions concern the notion of common Lyapunov function: when there exists a Lyapunov function that decays along all possible modes of the system (that is, for every possible constant value of the signal), then the system is uniformly asymptotically stable. Conversely, if the system is stable uniformly with respect to all signals switching in an arbitrary way, then a common Lyapunov function exists [75]. In the *linear* finite-dimensional case, the existence of a common Lyapunov function is actually equivalent to the global uniform exponential stability of the system [81] and, provided that the admissible modes are finitely many, the Lyapunov function can be taken polyhedral or polynomial [39], [40], [53]. A special role in the switched control literature has been played by common quadratic Lyapunov functions, since their existence can be tested rather efficiently (see [54] and references therein). Algebraic approaches to prove the stability of switched systems under arbitrary switching, not relying on Lyapunov techniques, have been proposed in [72], [26].

Other interesting issues concerning the stability of switched systems arise when, instead of considering arbitrary switching, one restricts the class of admissible signals, by imposing, for instance, a dwell time constraint [59].

Another rich area of research concerns discrete-time switched systems, where new intriguing phenomena appear, preventing the algebraic characterization of stability even for small dimensions of the state space [69]. It is known that, in this context, stability cannot be tested on periodic signals alone [42].

Finally, let us mention that little is known about infinite-dimensional switched system, with the exception of some results on uniform asymptotic stability ([78], [91], [92]) and some recent papers on optimal control ([58], [99]).

GEOMETRICA Project-Team

4. Application Domains

4.1. Geometric Modeling and Shape Reconstruction

Modeling 3D shapes is required for all visualization applications where interactivity is a key feature since the observer can change the viewpoint and get an immediate feedback. This interactivity enhances the descriptive power of the medium significantly. For example, visualization of complex molecules helps drug designers to understand their structure. Multimedia applications also involve interactive visualization and include e-commerce (companies can present their products realistically), 3D games, animation and special effects in motion pictures. The uses of geometric modeling also cover the spectrum of engineering, computer-aided design and manufacture applications (CAD/CAM). More and more stages of the industrial development and production pipeline are now performed by simulation, due to the increased performance of numerical simulation packages. Geometric modeling therefore plays an increasingly important role in this area. Another emerging application of geometric modeling with high impact is medical visualization and simulation.

In a broad sense, shape reconstruction consists of creating digital models of real objects from points. Example application areas where such a process is involved are Computer Aided Geometric Design (making a car model from a clay mockup), medical imaging (reconstructing an organ from medical data), geology (modeling underground strata from seismic data), or cultural heritage projects (making models of ancient and or fragile models or places). The availability of accurate and fast scanning devices has also made the reproduction of real objects more effective such that additional fields of applications are coming into reach. The members of GEOMETRICA have a long experience in shape reconstruction and contributed several original methods based upon the Delaunay and Voronoi diagrams.

4.2. Scientific Computing

Meshes are the basic tools for scientific computing using finite element methods. Unstructured meshes are used to discretize domains bounded by complex shapes while allowing local refinements. GEOMETRICA contributes to mesh generation of 2D and 3D possibly curved domains. Most of our methods are based upon Delaunay triangulations, Voronoi diagrams and their variants. Anisotropic meshes are also investigated. We investigate in parallel both greedy and variational mesh generation techniques. The greedy algorithms consist of inserting vertices in an initial coarse mesh using the Delaunay refinement paradigm, while the variational algorithms consists of minimizing an energy related to the shape and size of the elements. Our goal is to show the complementarity of these two paradigms. Quadrangle surface meshes are also of interest for reverse engineering and geometry processing applications. Our goal is to control the final edge alignment, the mesh sizing and the regularity of the quadrangle tiling.

GRACE Team

3. Application Domains

3.1. Cryptology

We want to establish the security of practical proposals relying on computational problems, be they standardized (like RSA or Elliptic Curve Cryptography), or more exotic (like Hyperelliptic Curve Cryptography). We do not work with abstract cryptographic primitives. On the design side, building efficient near-optimal codes impacts directly on the security of basic operations in symmetric primitives. We also investigate other applications, such as secret sharing schemes, universal hash functions, and message authentication, revisiting them in the context of Algebraic Geometry codes.

3.2. Codes in Computer Science

We do not want to do basic forward error correction, dealing with bit error rates and signal-to-noise ratios. Rather, we aim to deal with higher models of communication and computation, including peer-to-peer systems and distributed storage. We also consider adversarial noise, or distributed computations with byzantine faults. List decoding deals precisely with these kinds of "difficult", non-random errors. In a related spirit, one can deal with "computationally bounded channels", where the errors are generated by an adversarial machine or algorithm that is computationally bounded.

GRAND-LARGE Project-Team (section vide)

HIPERCOM Project-Team

4. Application Domains

4.1. Introduction

The HIPERCOM project-team is mainly concerned by six domains:

- wireless mobile ad hoc networks,
- services over mobile networks,
- community networks,
- vehicular networks,
- large ad hoc networks with sensor nodes,
- energy-efficient wireless sensor networks.

4.2. Wireless mobile ad hoc networks

Wireless mobile ad hoc networks, Services over mobile networks, Community Networks.

Abstract. Mobile wireless networks have numerous applications in rescue and emergency operation, military tactical networking and in wireless high speed access to the internet.

A mobile ad hoc network is a network made of a collection of mobile nodes that gather spontaneously and communicate without requiring a pre-existing infrastructure. Of course a mobile ad hoc network use a wireless communication medium. They can be applied in various contexts:

- military;
- rescue and emergency;
- high speed access to internet.

The military context is the most obvious application of mobile ad hoc networks.
Soldiers invading a country won't susbscribe in advance to the local operator. On the reverse side, home units won't use their local operators firstly because they will likely be disrupted in the first hours of the conflict, and secondly because a wireless communication via an operator is not stealth enough to protect the data and the units. In Checheny, a general has been killed by a missile tracking the uplink signal of his portable phone.
The rescue context is halfway between military and civilian applications. In the september 11 disaster, most of the phone base station of the area have knocked out in less than twenty minutes. The remaining base stations were unable to operate because they could not work in ad hoc mode. The Wireless Emergency Rescue Team recommanded afterward that telecom operators should provide ad hoc mode for their infrastructure in order to operate in emergency situation in plain cooperation with police, firemen and hospital networks.
Mobile ad hoc network provide an enhanced coverage for high speed wireless access to the internet. The now very popular WLAN standard, WiFi, provides much larger capacity than mobile operator networks. Using a mobile ad hoc network around hot spots will offer high speed access to much larger community, including cars, busses, trains and pedestrians.

4.3. Services over mobile networks

Abstract. New wireless network calls for new services that fullfil the requirement in terms of mobility and capacity.

The generalization of a new generation of mobile networks calls for a new set of services and applications. For example:

- Indoor and outdoor positioning
- Service discovery and localisation
- Multicast and quality of services

Quality of service has become the central requirement that users expect from a network. High throughput, service continuity are critical issue for multimedia application over the wireless internet where the bandwidth is more scarce than in the wired world. A significant issue in the ad-hoc domain is that of the integrity of the network itself. Routing protocols allow, according to their specifications, any node to participate in the network - the assumption being that all nodes are behaving well and welcome. If that assumption fails - then the network may be subject to malicious nodes, and the integrity of the network fails. An important security service over mobile networks is to ensure that the integrity of the network is preserved even when attacks are launched against the integrity of the network.

4.4. Community Networks

Abstract. There is an increasing demand to deploy network within a community, rural or urban, with cabled or wireless access.

Community networks or citizen network are now frequent in big cities. In America most of the main cities have a community network. A community network is using the communication resource of each member (ADSL, Cable and wireless) to provide a general coverage of a city. Pedestrian in the street or in city mails can communicate via a high speed mobile mesh network. This new trend now appears in Europe with many experiments of the OLSR routing protocol in Paris, Lille, Toulouse, Berlin, Bruxelles, Seattle. The management of such networks is completely distributed and makes them very robust to faults. There is room for smart operators in this business.

4.5. Vehicular Networks

Abstract. Intelligent transport systems require efficient wireless telecommunications.

Vehicular ad hoc networks (VANET) are based on short- to medium-range transmission systems that support both vehicle-to-vehicle and vehicle-to-roadside communications. Vehicular networks will enable vehicular safety applications (safety warnings) as well as non-safety applications (real-time traffic information, routing support, mobile entertainment, and many others). We are interested in developing an efficient routing protocol that takes advantage of the fixed network infrastructure deployed along the roads. We are also studying MAC layer issues in order to provide more priority for security messages which have stringent delivery constraints.

4.6. Large ad hoc networks with sensor nodes

Abstract. Large autonomous wireless sensors in the internet of the things need very well tuned algorithms.

Self-organization is considered as a key element in tomorrow's Internet architecture. A major challenge concerning the integration of self-organized networks in the Internet is the accomplishment of light weight network protocols in large ad hoc environments.

In this domain, Hipercom's activity with wireless sensor nodes in collaboration with the Freie Universitaet in Berlin explores various solutions, including extensions of OLSR (for example DHT-OLSR) using programmable sensor nodes co-designed by the Freie Universitaet, and provides one of the largest testbeds of this kind, to date.

4.7. Energy efficient wireless sensor networks

Abstract. Energy efficiency is a key property in wireless sensor networks.

Various techniques are used to contribute to energy efficiency. In the OCARI network, an industrial wireless sensor network, we have designed and implemented an energy efficient routing protocol and a node activity scheduling algorithm allowing router nodes to sleep. We have applied a cross-layering approach allowing the optimization of MAC and network protocols taking into account the application requirements and the environment in which the network operates. This activity has been done in collaboration with our partners EDF, LIMOS and TELIT.

IN-SITU Project-Team

4. Application Domains

4.1. Research Domains

INSITU works on general problems of interaction in multi-surface environments as well as on challenges associated with specific research groups. The former requires a combination of controlled experiments and field studies; the latter involves participatory design with users. We are currently working with highly creative people, particularly designers and music composers, to explore interaction techniques and technologies that support the earliest phases of the design process. We are also working with research scientists, particularly neuroscientists and astrophysicists, in our explorations of interaction in multisurface environments, and with doctors and nurses to support crisis management situations.

MACS Project-Team

4. Application Domains

4.1. Application domains

Our researches have natural applications in all sectors of the mechanical industry: car and naval industries; aeronautics and space; civil engineering; tires; MEMs and nanotechnologies...

We also actively seek new applications in biotechnologies, although of course the economy and structuring of this sector is not as developed yet.

MAXPLUS Project-Team

4. Application Domains

4.1. Systèmes à événements discrets (productique, réseaux)/Discrete event systems (manufacturing systems, networks)

Une partie importante des applications de l'algèbre max-plus provient des systèmes dynamiques à événements discrets [6]. Les systèmes linéaires max-plus, et plus généralement les systèmes dynamiques monotones contractants, fournissent des modèles naturels dont les résultats analytiques peuvent être appliqués aux problèmes d'évaluation de performance. Relèvent de l'approche max-plus, tout au moins sous forme simplifiée : des problèmes de calcul de temps de cycle pour des circuits digitaux [80], des problèmes de calcul de débit pour des ateliers [126], pour des réseaux ferroviaires [79] ou routiers, et l'évaluation de performance des réseaux de communication [71]. L'approche max-plus a été appliquée à l'analyse du comportement temporel de systèmes concurrents, et en particulier à l'analyse de "high level sequence message charts" [75], [134]. Le projet Maxplus collabore avec le projet Metalau, qui étudie particulièrement les applications des modèles max-plus à la modélisation microscopique du trafic routier [142], [139], [104].

English version

One important part of applications of max-plus algebra comes from discrete event dynamical systems [6]. Max-plus linear systems, and more generally, monotone nonexpansive dynamical systems, provide natural models for which many analytical results can be applied to performance evaluation problems. For instance, problems like computing the cycle time of asynchronous digital circuits [80], or computing the throughput of a workshop [126] or of a transportation network, and performance evaluation problems for communication networks, are often amenable to max-plus algebra, at least in some simplified form, see in particular [79] and [71]. The max-plus approach has been applied to the analysis of the time behaviour of concurrent systems, and in particular, to the analysis of high level sequence message charts [75], [134]. The Maxplus team collaborates with the Metalau team, working particularly on the applications of max-plus models to the microscopic modelling of road traffic [142], [139], [104].

4.2. Commande optimale et jeux/Optimal control and games

La commande optimale et la théorie des jeux ont de nombreuses applications bien répertoriées: économie, finance, gestion de stock, optimisation des réseaux, aide à la décision, etc. En particulier, le projet Mathfi travaille sur les applications à des problèmes de mathématiques financières. Il existe une tradition de collaborations entre les chercheurs des projets Mathfi et Maxplus sur ces questions, voir par exemple [5] qui comprend un résultat exploitant des idées de théorie spectrale non-linéaire, présentées dans [3].

English version

Optimal control and game theory have numerous well established applications fiels: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In particular, the Mathfi team works on applications in mathematical finance. There is a tradition of collaboration between researchers of the Maxplus team and of the Mathfi team on these questions, see as an illustration [5] where ideas from the spectral theory of monotone homogeneous maps [3] are applied.

4.3. Recherche opérationnelle/Operations research

L'algèbre max-plus intervient de plusieurs manières en Recherche opérationnelle. Premièrement, il existe des liens profonds entre l'algèbre max-plus et les problèmes d'optimisation discrète, voir [81]. Ces liens conduisent parfois à de nouveaux algorithmes pour les problèmes de recherche opérationnelle classiques, comme le problème de circuit de poids moyen maximum [88]. Certains problèmes combinatoires, comme des problèmes de programmation disjonctive, peuvent être décomposés par des méthodes de type max-plus [177]. Ensuite, le rôle de l'algèbre max-plus dans les problèmes d'ordonnancement est bien connu depuis les années 60, les dates de complétion pouvant souvent être calculées à partir d'équations linéaires max-plus. Plus récemment, des représentations de problèmes d'ordonnancement ont pu être obtenues à partir de semi-groupes de matrices max-plus : une première représentation a été obtenue dans [113] pour le cas du "jobshop", une représentation plus simple a été obtenue dans [136] dans le cas du "flowshop". Ce point de vue algébrique a été très utile dans le cas du "flowshop" : il permet de retrouver des résultats anciens de dominance et d'obtenir ainsi de nouvelles bornes [136]. Finalement, en regardant l'algèbre max-plus comme une limite de l'algèbre classique, on peut utiliser des outils algébriques en optimisation combinatoire [132].

English version

Max-plus algebra arise in several ways in Operations Research. First, there are intimate relations between max-plus algebra and discrete optimisation problems, see [81]. Sometimes, these relations lead to new algorithms for classical Operations Research problems, like the maximal circuit mean [88]. There are also special combinatorial problems, like certain problems of disjunctive programming, which can be decomposed by max-plus type methods [177]. Next, the role of max-plus algebra in scheduling problems has been known since the sixties: completion dates can often be computed by max-plus linear equations. Recently, representations of certain scheduling problems using max-plus matrix semigroups have appeared, a first representation was given in [113] for the jobshop case, a simpler representation was given in [136] in the flowshop case. This algebraic point of view turned out to be particularly fruitful in the flowshop case: it allows one to recover old dominance results and to obtain new bounds [136]. Finally, viewing max-plus algebra as a limit of classical algebra allows to use algebraic tools in combinatorial optimisation [132].

4.4. Analyse statique de programmes/Static analysis of computer programs

L'interprétation abstraite est une technique, introduite par P. et R. Cousot [92], qui permet de déterminer des invariants de programmes en calculant des points fixes minimaux d'applications monotones définies sur certains treillis. On associe en effet à chaque point de contrôle du programme un élément du treillis, qui représente une sur-approximation valide de l'ensemble des valeurs pouvant être prises par les variables du programme en ce point. Le treillis le plus simple exprimant des propriétés numériques est celui des produits Cartésiens d'intervalles. Des treillis plus riches permettent de mieux tenir compte de relations entre variables, en particulier, des classes particulières de polyèdres sont souvent employées.

Voici, en guise d'illustration, un petit exemple de programme, avec le système de point fixe associé, pour le treillis des intervalles:

```
   void main() {
int x=0;          // 1
while (x<100) {   // 2
  x=x+1;          // 3
}                 // 4
         }
```

$$\begin{aligned} x_1 &= [0,0] \\ x_2 &=]-\infty,99] \cap (x_1 \cup x_3) \\ x_3 &= x_2 + [1,1] \\ x_4 &= [100,+\infty[\cap (x_1 \cup x_3) \end{aligned}$$

Si l'on s'intéresse par exemple aux valeurs maximales prise par la variable x au point de contrôle 2, soit $x_2^+ := \max x_2$, après une élimination, on parvient au problème de point fixe:

$$x_2^+ = \min\left(99, \max\left(0, x_2^+ + 1\right)\right) , \tag{1}$$

qui a pour plus petite solution $x_2^+ = 99$, ce qui prouve que x est majoré par 99 au point 2.

On reconnait ici un opérateur de point fixe associé à un problème de jeux à deux joueurs et somme nulle. Cette analogie est en fait générale, dans le cadre d'un collaboration que l'équipe entretient depuis plusieurs années avec l'équipe MeASI d'Eric Goubault (CEA et LIX), spécialiste d'analyse statique, nous avons en effet mis progressivement en évidence une correspondance [91], [110], entre les problèmes de jeux à somme nulle et les problèmes d'analyse statique, qui peut se résumer par le dictionnaire suivant:

Jeux	Interprétation abstraite
système dynamique	programme
opérateur de Shapley	fonctionnelle
espace d'état	(# points de contrôle) $\times$ (# degrés de liberté du treillis)
problème en horizon n	exécution de n pas
limite du problème en horizon fini	invariant optimal (borne)
itération sur les valeurs	itération de Kleene

Pour que le nombre d'états du jeu soit fini, il est nécessaire de se limiter à des treillis d'ensembles ayant un nombre fini de degrés de liberté, ce qui est le cas de domaines communément utilisés (intervalles, ensembles définis par des contraintes de potentiel de type $x_i - x_j \leq$ cst, mais aussi, les "templates" qui sont des sous-classes de polyèdres introduits récemment par Sankaranarayanan, Sipma et Manna [166]). L'ensemble des actions est alors fini si on se limite à une arithmétique affine. Signalons cependant qu'en toute généralité, on aboutit à des jeux avec un taux d'escompte négatif, ce qui pose des difficultés inédites. Cette correspondance entre jeux et analyse statique est non intuitive, au sens où les actions du minimiseur consistent à sélectionner des points extrêmes de certains polyèdres obtenus par un mécanisme de dualité.

Une pathologie bien répertoriée en analyse statique est la lenteur des algorithmes de point fixe, qui peuvent effectuer un nombre d'itérations considérable (99 itérations pour obtenir le plus petit point fixe de (8)). Celle-ci est usuellement traitée par des méthodes d'accélération de convergence dites d'élargissement et rétrécissement [93], qui ont cependant l'inconvénient de conduire à une perte de précision des invariants obtenus. Nous avons exploité la correspondance entre analyse statique et jeux pour développer des algorithmes d'une nature très différente, s'inspirant de nos travaux antérieurs sur l'itération sur les politiques pour les jeux répétés [111], [86], [87],[7]. Une version assez générale de cet algorithme, adaptée au domaine des templates, est décrite dans [110] et a fait l'objet d'une implémentation prototype. Chaque itération combine de la programmation linéaire et des algorithmes de graphes. Des résultats expérimentaux ont montré le caractère effectif de la méthode, avec souvent un gain en précision par rapport aux approches classiques, par exemple pour des programmes comprenant des boucles imbriquées.

Ce domaine se trouve être en pleine évolution, un enjeu actuel étant de traiter d'une manière qui passe à l'échelle des invariants plus précis, y compris dans des situations où l'arithmétique n'est plus affine.

English version

The abstract interpretation method introduced by P. and R. Cousot [92], allows one to determine automatically invariants of programs by computing the minimal fixed point of an order preserving map defined on a complete lattice. To every breakpoint of the program is associated an element of the lattice, which yields a valid overapproximation of the set of reachable values of the vectors of variables of the program, at this breakpoint. The simplest lattice expressing numerical invariants consists of Cartesian products of intervals. More sophisticated lattices, taking into account relations between variables, consisting in particular of subclasses of polyhedra, are often used.

As an illustration, we gave before Eqn (8) a simple example of program, together with the associated fixed-point equation. In this example, the value of the variable x at the breakpoint 2 is bounded by the smallest solution x_2^+ of the fixed point problem (8), which is equal to 99.

The fixed point equation (8) is similar to the one arising in the theory of zero-sum repeated games. This analogy turns out to be general. Un a series of joint works of our team with the MeASI team of Eric Goubault (CEA and LIX), we brought progressively to light a correspondence [91], [110], between the zero-sum game problems and the static analysis problems, which can be summarized by the following dictionnary:

Games	Abstract interpretation
dynamical system	program
Shapley operator	functional
state space	(# breakpoints) $\times$ (# degrees of freedom)
horizon n problem	execution of n logical steps
limit of the value in horizon n	optimal invariant (bound)
value iteration	Kleene iteration

For the game to have a finite state space, we must restrict our attention to lattices of sets with a finite number of degrees of freedom, which is the case of the domains commonly used in static analysis (intervals, sets defined by potentials constraints of the form $x_i - x_j \leq$ cst, and also the subclasses of polyhedra called "templates", introduced recently by Sankaranarayanan, Sipma and Manna [166]). Then, the action space is finite if the arithmetics of the program is affine. However, in full generality, the games we end up with have a negative discount rate, which raises difficulties which are unfamiliar from the game theory point of view. This correspondence between games and static analysis turns out to be non intuitive, in that the action of the minimizer consist of selecting an extreme point of a polyhedron arising from a certain duality construction.

A well known pathology in static analysis is the fact that the standard Kleene fixed point algorithm may have a very slow behavior (99 iterations are needed to get the smallest fixed point of (8)). This is usually solved by using some accelerations of convergence, called widening and narrowing [93], which however lead to a loss of precision. We exploited the correspondence between static analysis and games to develop algorithms of a very different nature, inspired by our earlier work on policy iteration for games [111], [86], [87],[7]. A rather general version of this policy iteration algorithm, adpated to the domain of templates, is described in [110], together with a prototype implementation. Every iteration combines linear programming and combinatorial algorithms. Some experimental results indicate that the method often leads to invariants which are more accurate than the ones obtained by alternative methods, in particular for some programs with nested loops.

This topic of research is currently evolving, a question of current interest being to find accurate invariants, in a scalable way, in situations in which the arithmetics is not affine.

4.5. Autres applications/Other applications

L'algèbre max-plus apparaît de manière naturelle dans le calcul de scores de similitudes dans la comparaison de séquences génétiques. Voir par exemple [90].

English version

Max-plus algebra arises naturally in the computation of similarity scores, in biological sequence comparison. See for instance [90].

MEXICO Project-Team

4. Application Domains

4.1. Panorama

telecommunications, multimedia, transportation systems, web services

MExICo's research is motivated by problems on system management in several domains:

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize *adaptators* for every component in order to steer its interaction with the surrounding processes.
- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

This list is likely to grow over the next years as we continue our research.

4.2. Autonomous Telecommunications Systems: In-Band Supervision

Participants: Stefan Haar, Serge Haddad.

In the context of traditional hard-wired communication networks, supervision structures for managing faults, configuration, provisioning etc could be developed with a fixed infrastructure, and perform the communication between sensors, supervisors, policy enforcement points etc over a separate network using separate hardware. This rigid, **out-of-band** technology does not survive passing to today's and tomorrow's services and networks. In fact, the dynamic mobility of services combined across sites and domains cannot be captured unless the network used for supervision evolves in the same way and simultaneously, which rules out static solutions; but providing out-of-band infrastructure that grows with the networks to be supervised would be prohibitively expensive, if at all technically feasible. *Heterogeneity* is the other feature of modern networks that forces a change, since different domains are not likely to agree on a pervasive third-party supervision. Rather, the providers will keep control over the internal state and evolution of their domain, and accept only exchange through standardized outward interfaces.

Supervision has thus to be re-invented on an *in-band*, *autonomous* base: monitoring probes deployed on the web, dysfunctions on one peer node diagnosed by another peer in a network with changing configuration, enhanced supervisor and actor capacities of services, etc. *MExICo* will work on improving the interoperability of service components through continued application of e.g. distributed techniques for control and diagnosis.

4.3. Web Services

Participants: Stefan Haar, Serge Haddad.

Specific applications targeted by *MExICo* include the problem of adaptation in Service-Oriented Computing (SOC). The challenge is here twofold, stemming both from the distributed nature of services (scattered over the entire web) and their heterogeneous origins.

4.3.1. Context

Web services have become the most frequently used model of design and programming based on components for business applications. Web service languages like BPEL have useful constructors that manage for instance exceptions, (timed guarded) waiting of messages, parallel execution of processes, distant service invocations, etc. Interoperability of components is based on interaction protocols associated with them and often published on public or private registers. In the framework of Web services, these protocols are called abstract processes by contrast with business processes (i.e. services). Composition of components must be analyzed for several reasons and at least to avoid deadlocks during execution. This has led to numerous works that focus on compositional verification, substitution of a component by another one, synthesis of adaptators, etc., and triggered a push towards a unifying theoretical framework (see e.g. [115], [119])

4.3.2. Problems

Interoperability requires that when a user or a program wants to interact with the component, the knowledge of the interaction protocol is enough. Our previous works have shown that the interaction protocols can be inherently ambiguous: no client can conduct a correct interaction with the component in every scenario. This problem is even more complex when the protocol can evolve during execution due to adaptation requirements. The composition of components also raises interesting problems. When composing optimal components (w.r.t. the number of states for instance) the global component can be non optimal. So one aims at reducing a posteriori or better on the fly the global component. At last, the dynamical insertion of a component in a business process requires to check whether this insertion is behaviorally consistent [121], [109]

We do not intend to check global properties based on a modular verification technique. Rather, given an interaction protocol per component and a global property to ensure, we want to synthesize an adaptator per component such that this property is fulfilled or to detect that there cannot exist such adaptators [106]. In another research direction, one can introduce the concept of utility of a service and then optimize a system i.e. keeping the same utility value while reducing the resources (states, transitions, clocks, etc.).

OAK Team

4. Application Domains

4.1. Online content management

This concerns archiving filtered content from online information sources (journals, blogs, ...) with the purpose of recording their perspective on facts involving specific countries, regions or enterprises, key actors etc. We have recently explored such an application in a demonstration proposal, built using our XR model for XML documents with semantic annotations (currently under evaluation). The benefit of XR was to simplify and streamline the specification of document- and semantics-rich data management applications. Support for documents allows to keep an evidence (source) for facts and statements that may be extracted from them. At the same time, support for semantics allows to connect documents to people, ideas, trends etc. and to reason across data sources [9].

4.2. Open data intelligence

Open data intelligence: the goal is to build and efficiently exploit warehouses of Open Data, integrated from several data sources on the Web, in order to produce consolidated rich information to be given to decision makers and to the citizens. Such projects have been started in France notably by the city areas of Rennes (pioneer in Open Data usage), Paris, and more recently by Grenoble, in a project to which we participate; the ICT Labs DataBridges also focused on such topics. Oak competencies required for such projects are related to large-scale RDF data management as well as to the design of innovative data models for semantic-rich content.

4.3. Efficient complex data management in the cloud

In a cloud environment, data catalogs and indices need to be efficiently built and maintained, typically in parallel; queries need to be routed to only those data subsets which are likely to lead to results, and efficiently executed, using parallelism and the available indexes. We work on such topics within the Europa ICT Labs activity. Algorithms for efficiently handling indexes and views in the cloud for heterogeneous, complex-structure data are also at the core of our work proposal in the Datalyse project (see below). Our 2012 work in this context has lead to [10], [12], [13].

4.4. 360 degree customer view

Companies seek to gather as much information as possible about their customers, for instance for more targeted publicity campaigns, market analysis, offer personalization, etc. That is, they want to consider data beyond the data they collected about their customers in their proprietary databases. Complementary data include for instance social data extracted from customers' activities in social networks, public data related to their place of residence (e.g., crime rate or housing price evolution). To achieve this goal, data integration on highly heterogeneous and massive data is necessary. Furthermore, as one means to assess both the correctness of the integration result and the quality (in this application most notably trustworthiness), we can resort to data provenance. These topics are explored in the project Datalyse which we submitted in 2012 to the French national "AAP Cloud 3: Big Data" call, a project headed by the "Business & Decision" company and whose ongoing evaluation will continue in early 2013.

PARIETAL Project-Team

4. Application Domains

4.1. Inverse problems in Neuroimaging

Many problems in neuroimaging can be framed as forward and inverse problems. For instance, the neuroimaging *inverse problem* consists in predicting individual information (behavior, phenotype) from neuroimaging data, while an important the *forward problem* consists in fitting neuroimaging data with high-dimensional (e.g. genetic) variables. Solving these problems entails the definition of two terms: a loss that quantifies the goodness of fit of the solution (does the model explain the data reasonably well ?), and a regularization schemes that represents a prior on the expected solution of the problem. In particular some priors enforce some properties of the solutions, such as sparsity, smoothness or being piecewise constant.

Let us detail the model used in the inverse problem: Let X be a neuroimaging dataset as an (n_{subj}, n_{voxels}) matrix, where n_{subj} and n_{voxels} are the number of subjects under study, and the image size respectively, Y an array of values that represent characteristics of interest in the observed population, written as (n_{subj}, n_f) matrix, where n_f is the number of characteristics that are tested, and β an array of shape (n_{voxels}, n_f) that represents a set of pattern-specific maps. In the first place, we may consider the columns $Y_1, .., Y_{n_f}$ of Y independently, yielding n_f problems to be solved in parallel:

$$Y_i = X\beta_i + \epsilon_i, \forall i \in \{1, .., n_f\},$$

where the vector contains β_i is the i^{th} line of β. As the problem is clearly ill-posed, it is naturally handled in a regularized regression framework:

$$\widehat{\beta}_i = \text{argmin}_{\beta_i} \|Y_i - X\beta_i\|^2 + \Psi(\beta_i), \quad (2)$$

where Ψ is an adequate penalization used to regularize the solution:

$$\Psi(\beta; \lambda_1, \lambda_2, \eta_1, \eta_2) = \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 + \eta_1 \|\nabla\beta\|_1 + \eta_2 \|\nabla\beta\|_2^2 \quad (3)$$

with λ_1, λ_2, η_1, $\eta_2 \geq 0$. In general, only one or two of these constraints will have a non-zero weighting:

- When $\lambda_1 > 0$ only (LASSO), and to some extent, when $\lambda_1, \lambda_2 > 0$ only (elastic net), the optimal solution β is (possibly very) sparse, but may not exhibit a proper image structure; it does not fit well with the intuitive concept of a brain map.
- Total Variation regularization (see Fig. 1) is obtained for ($\eta_1 > 0$ only), and typically yields a piecewise constant solution.
- Smooth lasso is obtained with ($\eta_2 > 0$ and $\lambda_1 > 0$ only), and yields smooth, compactly supported spatial basis functions.

The performance of the predictive model can simply be evaluated as the amount of variance in Y_i fitted by the model, for each $i \in \{1, .., n_f\}$. This can be computed through cross-validation, by *learning*$\widehat{\beta}_i$ on some part of the dataset, and then estimating $(Y_i - X\widehat{\beta}_i)$ using the remainder of the dataset.

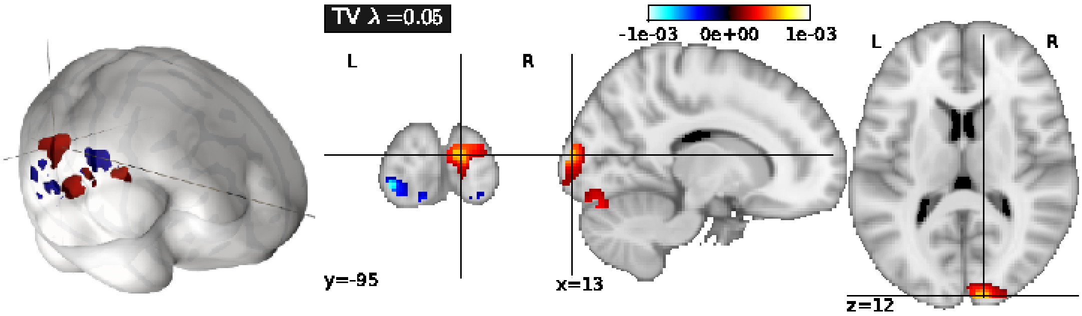

Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problems here consists in predicting the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the observation of an image. Learning and test are performed across individuals. Unlike other approaches, Total Variation regularization yields a sparse and well-localized solution that enjoys particularly high accuracy.

This framework is easily extended by considering

- *Grouped penalization*, where the penalization explicitly includes a prior clustering of the features, i.e. voxel-related signals, into given groups. This is particularly important to include external anatomical priors on the relevant solution.
- *Combined penalizations*, i.e. a mixture of simple and group-wise penalizations, that allow some variability to fit the data in different populations of subjects, while keeping some common constraints.
- *Logistic regression*, where a sigmoid non-linearity is applied to the linear model so that it yields a probability of classification in a binary classification problem.
- *Robustness to between-subject variability* is an important question, as it makes little sense that a learned model depends dramatically on the particular observations used for learning. This is an important issue, as this kind of robustness is somewhat opposite to sparsity requirements.
- *Multi-task learning*: if several target variables are thought to be related, it might be useful to constrain the estimated parameter vector β to have a shared support across all these variables.
  For instance, when one of the variables Y_i is not well fitted by the model, the estimation of other variables $Y_j, j \neq i$ may provide constraints on the support of β_i and thus, improve the prediction of Y_i. Yet this does not impose constraints on the non-zero parameters of the parameters β_i.

$$Y = X\beta + \epsilon, \tag{4}$$

then

$$\widehat{\beta} = \text{argmin}_{\beta=(\beta_i),i=1..n_f} \sum_{i=1}^{n_f} \|Y_i - X\beta_i\|^2 + \lambda \sum_{j=1}^{n_{voxels}} \sqrt{\sum_{i=1}^{n_f} \beta_{i,j}^2} \tag{5}$$

4.2. Multivariate decompositions

Multivariate decompositions are an important tool to model complex data such as brain activation images: for instance, one might be interested in extracting an atlas of brain regions from a given dataset, such as regions depicting similar activities during a protocol, across multiple protocols, or even in the absence of protocol (during resting-state). These data can often be factorized into spatial-temporal components, and thus can be estimated through *regularized Principal Components Analysis* (PCA) algorithms, which share some common steps with regularized regression.

Let X be a neuroimaging dataset written as an (n_{subj}, n_{voxels}) matrix, after proper centering; the model reads

$$X = AD + \epsilon, \tag{6}$$

where D represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}), and A the associated subject-wise loadings. While traditional PCA and independent components analysis are limited to reconstruct components D within the space spanned by the column of X, it seems desirable to add some constraints on the rows of D, that represent spatial maps, such as sparsity, and/or smoothness, as it makes the interpretation of these maps clearer in the context of neuroimaging.

This yields the following estimation problem:

$$\min_{D,A} \|X - AD\|^2 + \Psi(D) \text{ s.t. } \|A_i\| = 1 \ \forall i \in \{1..n_f\}, \tag{7}$$

where $(A_i),\ i \in \{1..n_f\}$ represents the columns of A. Ψ can be chosen such as in Eq. (2) in order to enforce smoothness and/or sparsity constraints.

The problem is not jointly convex in all the variables but each penalization given in Eq (2) yields a convex problem on D for A fixed, and conversely. This readily suggests an alternate optimization scheme, where D and A are estimated in turn, until convergence to a local optimum of the criterion. As in PCA, the extracted components can be ranked according to the amount of fitted variance. Importantly, also, estimated PCA models can be interpreted as a probabilistic model of the data, assuming a high-dimensional Gaussian distribution (probabilistic PCA).

4.3. Covariance estimation

Another important estimation problem stems from the general issue of learning the relationship between sets of variables, in particular their covariance. Covariance learning is essential to model the dependence of these variables when they are used in a multivariate model, for instance to assess whether an observation is aberrant or not or in classification problems. Covariance learning is necessary to model latent interactions in high-dimensional observation spaces, e.g. when considering multiple contrasts or functional connectivity data.

The difficulties are two-fold: on the one hand, there is a shortage of data to learn a good covariance model from an individual subject, and on the other hand, subject-to-subject variability poses a serious challenge to the use of multi-subject data. While the covariance structure may vary from population to population, or depending on the input data (activation versus spontaneous activity), assuming some shared structure across problems, such as their sparsity pattern, is important in order to obtain correct estimates from noisy data. Some of the most important models are:

- **Sparse Gaussian graphical models**, as they express meaningful conditional independence relationships between regions, and do improve conditioning/avoid overfit.
- **Decomposable models**, as they enjoy good computational properties and enable intuitive interpretations of the network structure. Whether they can faithfully or not represent brain networks is an important question that needs to be addressed.
- **PCA-based regularization of covariance** which is powerful when modes of variation are more important than conditional independence relationships.

Adequate model selection procedures are necessary to achieve the right level of sparsity or regularization in covariance estimation; the natural evaluation metric here is the out-of-samples likelihood of the associated Gaussian model. Another essential remaining issue is to develop an adequate statistical framework to test differences between covariance models in different populations. To do so, we will consider different means of parametrizing covariance distributions and their impact on the network. Our current work on post-stroke patients (see e.g. Fig. 2) suggests indeed that modeling may prove essential to perform sensitive inference.

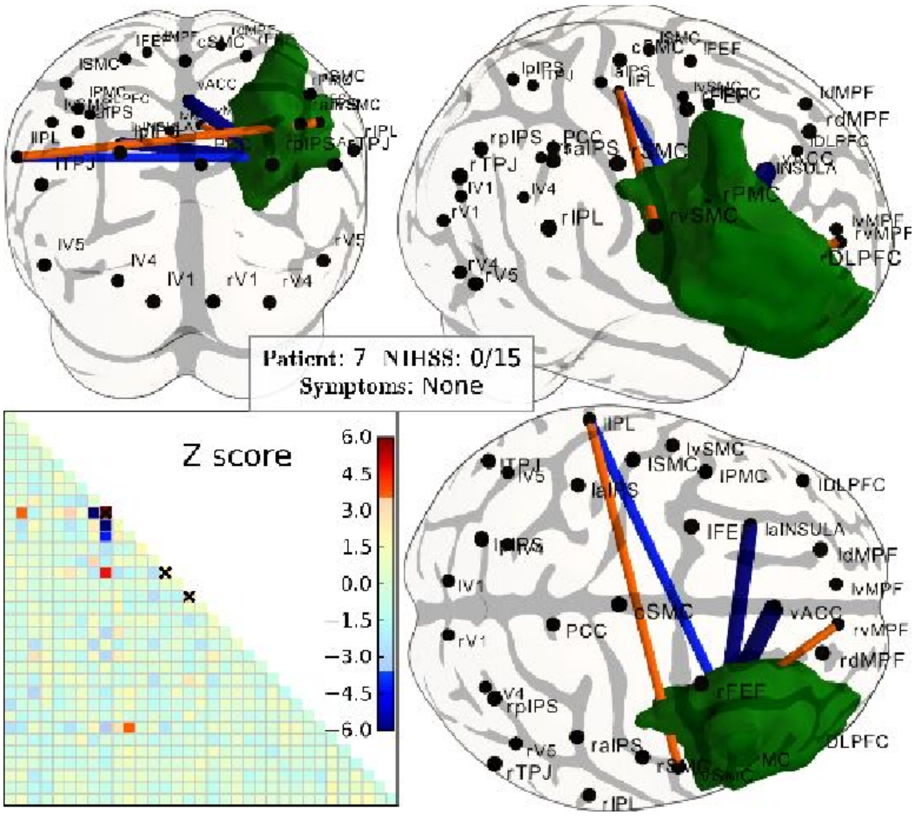

Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion outlined in green) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical detection of the difference requires a sophisticated statistical framework for the comparison of graphical models.

PARSIFAL Project-Team

4. Application Domains

4.1. Automated Theorem Proving

Automated theorem proving has traditionally focused on classical first-order logic, but non-classical logics are increasingly becoming important in the specification and analysis of software. Most type systems are based on (possibly second-order) propositional intuitionistic logic, for example, while resource-sensitive and concurrent systems are most naturally expressed in linear logic.

The members of the Parsifal team have a strong expertise in the design and implementation of performant automated reasoning systems for such non-classical logics. In particular, the Linprover suite of provers [35] continue to be the fastest automated theorem provers for propositional and first-order linear logic.

Any non-trivial specification, of course, will involve theorems that are simply too complicated to prove automatically. It is therefore important to design semi-automated systems that allow the user to give high level guidance, while at the same time not having to write every detail of the formal proofs. High level proof languages in fact serve a dual function – they are more readily comprehended by human readers, and they tend to be more robust with respect to maintenance and continued evolution of the systems. Members of the Parsifal team, in association with other Inria teams and Microsoft Research, have been building a heterogeneous semi-automatic proof system for verifying distributed algorithms [36].

On a more foundational level, the team has been developing many new insights into the structure of proofs and the proof search spaces. Two directions, in particular, present tantalizing possibilities:

- The concept of *multi-focusing* [37] can be used to expose concurrency in computational behavior, which can in turn be exploited to prune areas of the proof search space that explore irrelevant interleavings of concurrent actions.
- The use of *bounded search*, where the bounds can be shown to be complete by meta-theoretic analysis, can be used to circumvent much of the non-determinism inherent in resource-sensitive logics such as linear logic. The lack of proofs of a certain bound can then be used to justify the presence or absence of properties of the encoded computations.

Much of the theoretical work on automated reasoning has been motivated by examples and implementations, and the Parsifal team intends to continue to devote significant effort in these directions.

4.2. Mechanized Metatheory

There has been increasing interest in the use of formal methods to provide proofs of properties of programs and programming languages. Tony Hoare's Grand Challenge titled "Verified Software: Theories, Tools, Experiments" has as a goal the construction of "verifying compilers" for a world where programs would only be produced with machine-verified guarantees of adherence to specified behavior. Guarantees could be given in a number of ways: proof certificates being one possibility.

The POPLMark challenge [33] envisions "a world in which mechanically verified software is commonplace: a world in which theorem proving technology is used routinely by both software developers and programming language researchers alike." The proposers of this challenge go on to say that a "crucial step towards achieving these goals is mechanized reasoning about language metatheory."

The Parsifal team has developed several tools and techniques for reasoning about the meta-theory of programming languages. One of the most important requirements for programming languages is the ability to reason about data structures with binding constructs up to α-equivalence. The use of higher-order syntax and nominal techniques for such data structures was pioneered by Miller, Nadathur and Tiu. The Abella system (see Section 3.2) implements a refinement of a number of these ideas and has been used to give full solutions to sections of the POPLMark challenge in addition to fully formal proofs of a number of other theorems in the meta-theory of the λ-calculus.

Now that the Abella system has been in circulation among colleagues during the past couple of years, there are many aspects of the methodology that now need to be addressed. During the summer of 2011, the team employed three interns Carnegie Mellon University and McGill University to work on different aspects of Abella. Particular focus was given to better ways to manipulate specification-logic contexts in the reasoning-logic and with finding ways to have Abella output a proper proof object (different from the scripts that are used to find a proof).

Our colleague Alwen Tiu from the Australian National University has also been building on our Bedwyr model checking tool so that we can build on top of it his SPEC system for doing model checking of spi-calculus expressions. We have adopted his enhancements to Bedwyr and are developing further improvements within the context of the BATT project (see Section 5.2).

4.3. Proof Certificates

Members of the Parsifal team have shown how to specify a large variety of proof systems—including natural deduction, the sequent calculus, and various tableau and free deduction systems—uniformly using either focused linear logic [52], [50] or focused intuitionistic logic [43] as the meta-language. In the presence of induction and co-induction, arbitrary finite computations can be embedded into single synthetic steps [34]. Additional work [8] shows that this same framework can also capture resolution refutations as well as Pratt primality certificates.

An important application then of this work in designing synthetic inference systems based on classical and intuitionistic logic is that of designing a *broad spectrum proof certificate*. The definition of proof certificates can be remarkably flexible within the simple setting of focused proofs.

The most important implications of such a certificate format would be that most of the worlds theorem provers should be able to print out their proofs and communication them to other provers: these other provers could then check such certificates by expanding the synthetic connectives they contain down into a small and fixed set of “micro” inference rules.

4.4. Automated reasoning and SMT solving

Automated reasoning uses a broad range of techniques whose soundness and completeness relate to the existence of proofs. The research programme of the ANR PSI project at Parsifal is to build a finer-grained connection by specifying automated reasoning techniques as the step-by-step construction of proofs, as we know it from proof theory and logic programming. The goal is to do this in a unifying framework, namely proof-search in a polarized and focussed logic. One of the advantages of this is to combine those techniques more easily. Another one is to envisage extending those techniques.

For instance, SAT-modulo-Theory problems require the combination of logical reasoning with domain-specific decision procedures. So in the PSI project we study how to incorporate the call to decision procedures in proof-theoretical framework like the focussed sequent calculus, and the proof-search mechanisms that are related to it.

In the same spirit we also study how to handle existential variables and equality, for which specific automated reasoning techniques have been designed (superposition / paramodulation calculi).

POEMS Project-Team

4. Application Domains

4.1. Introduction

We are concerned with all application domains where linear wave problems arise: acoustics and elastodynamics (including fluid-structure interactions), electromagnetism and optics, and gravity water waves. We give in the sequel some details on each domain, pointing out our main motivations and collaborations.

4.2. Acoustics

As the acoustic propagation in a fluid at rest can be described by a scalar equation, it is generally considered by applied mathematicians as a simple preliminary step for more complicated (vectorial) models. However, several difficult questions concerning coupling problems have occupied our attention recently. Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, is for our team a new and very challenging topic, which gives rise to a lot of open questions, from the modeling (Euler equations, Galbrun equations) to the numerical approximation of such models (which poses new difficulties). Our works in this area are partially supported by EADS and Airbus. The typical objective is to reduce the noise radiated by Airbus planes. Vibroacoustics, which concerns the interaction between sound propagation and vibrations of thin structures, also raises up a lot of relevant research subjects.

Both applications (Aeroacoustics and Vibroacoustics) led us in particular to develop an academic research between volumis methods and integral equations in time domain.

Finally, A particularly attractive application concerns the simulation of musical instruments, whose objectives are both a better understanding of the behavior of existing instruments and an aid for the manufacturing of new instruments. The modeling and simulation of the timpani and of the guitar have been carried out in collaboration with A. Chaigne of ENSTA. We are currently on the piano.

4.3. Electromagnetism

This is a particularly important domain, first because of the very important technological applications but also because the treatment of Maxwell's equations is much more technically involved from the mathematical point of view that the scalar wave equation. Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology, electromagnetic compatibility, design of optoelectronic micro-components or smart materials. Stealth technology relies in particular on the conception and simulation of new absorbing materials (anisotropic, chiral, non-linear...). The simulation of antennas raises delicate questions related to the complexity of the geometry (in particular the presence of edges and corners). In optics, the development of the Mmcro and nano optics has made recently fantastic progress and the thematic of metamaterials (with negative index of refraction) opens new amazing applications. For all these reasons, we are developing an intense research in the following areas

- Highly accurate and hybrid numerical methods in collaboration with CEA (Gramat) and ONERA (Toulouse).
- Electromagmetic wave propagation in periodic media.
- Development of simplified approximate models by asymptotic anslysis for various applications : boundary layers, thin coatings, thin domains, thin wires and cables, ...
- Mathematical and numerical questions linked to the modeling of metamaterials.

4.4. Elastodynamics

Wave propagation in solids is with no doubt, among the three fundamental domains that are acoustics, electromagnetism and elastodynamics, the one that poses the most significant difficulties from mathematical and numerical points of view.

Our activity on this topic began with applications in geophysics, which unfortunately has been forced to slow down in the middle of the 90's due to the disengagement of French oil companies in matter of research. However it has seen a most welcomed rebound through new academic problems (in parficular surface waves, perfectly matched layers techniques, inverse problems in wave guides) and industrial contacts, more precisely with CEA-LIST with which we have developed a long term collaboration in the domain of non destructive testing by ultrasounds. The most recent problems we have been dealing with in this domain concern elastic wave propagation in plates, the modeling of piezoelectric devices or elastic wave propagation in highly heterogeneous media.

POPIX Exploratory Action

4. Application Domains

4.1. Pharmacometrics

Participants: Marc Lavielle, Kevin Bleakley, Maud Delattre, Cyprien Mbogning, Célia Barthelemy, Hector Mesa, Elodie Maillot, Elena Carvajal, Laura Brocco.

POPIX is directly implicated in the domain of pharmacology. Historically, Marc Lavielle was the driving force behind the pharmacological modeling software MONOLIX, now an industry standard. Lixoft, an Inria start-up, now develops and supports MONOLIX and the commercial side of things. POPIX collaborates closely with Lixoft to transfer research results into software improvements and the development of new user tools in MONOLIX.

POPIX is also majorly implicated in the 5-year DDMoRe (Drug and Disease Model Resources) European project financed by the IMI (Innovative Medicines Initiative), a public-private partnership. In particular, POPIX has the task of developing new tools and methods for this project regrouping researchers in pharmacometrics, biostatistics and biology from both the public and private sectors. Specific tools and methods being developed by POPIX include:

- a clinical trial simulator (see the Software section for more details)
- protocol optimization tools
- diagnostic tools
- model selection tools
- estimation techniques for complex models (eg, stochastic differential equations, partial differential equations)

4.2. Pharmacogenetics

Participants: Marc Lavielle, Kevin Bleakley.

Medicine, even when prescribed following dosage rules, is an important cause of illness and death. In essence, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Pharmacogenetics, a subdomain of pharmacology, is the study of the the relationship between genetic variability and the therapeutic outcome. The future goal is "personal medicine" whereby the drug and dose are chosen with respect to the individual's genetic make-up.

Currently, in the population approach followed by POPIX, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much more harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient. The subsequent model-selection problem is thus very complicated, and requires powerful classification and statistical learning tools.

4.3. Oncology

Participants: Marc Lavielle, Célia Barthelemy.

Despite great advances in the treatment and diagnosis of cancer, many steps remain to further improve prognoses and quality of life of cancer patients. Numerical models can be used to help adapt treatment protocol to the characteristics of each patient, ie, improve treatment efficacy by:

- choosing the best treatment
- choosing the best dose
- choosing the best drug-delivery protocol
- optimizing the above parameters to minimize toxicity

POPIX is part of the Inria pluridiciplinary project MONICA (MOdèles Numériques et Imagerie pour le CAncer), including the NUMED, MC2 and ASCLEPIOS Inria teams, that aims to optimize the parameters listed above using numerical modeling.

Collaborations with NUMED and MC2 are already underway, in particular with the aim of extending the statistical methods developed by POPIX to partial differential equations based models. NUMED works on models of tumor growth and has already implemented an extension of MONOLIX to KPP-type reaction-diffusion models.

4.4. Change-point detection in signals

Participants: Marc Lavielle, Kevin Bleakley.

Change-point detection is historically a signal processing problem whereby we search for points at which a 1-dimensional noisy signal has an abrupt change in some way, eg. change in mean or variance. It turns out that similar methods can be developed for finding the genomic locations at which the DNA copy number changes in a cancer-stricken (or other) patient. Normally, we have two copies of DNA along the whole genome, so specific changes (gains or losses) in copy number can be associated with the specific cancer, hopefully leading to treatment possibilites. Kevin Bleakley collaborates with researchers at the Curie Institute in change-point detection in one or many DNA copy number profiles. Related to this, Marc Lavielle and Kevin Bleakley have developed methods to find change-points in histogram probability density functions using data sampled from the (unknown) density.

REGULARITY Project-Team

4. Application Domains

4.1. Uncertainties management

Our theoretical works are motivated by and find natural applications to real-world problems in a general frame generally referred to as uncertainty management, that we describe now.

Since a few decades, modeling has gained an increasing part in complex systems design in various fields of industry such as automobile, aeronautics, energy, etc. Industrial design involves several levels of modeling: from behavioural models in preliminary design to finite-elements models aiming at representing sharply physical phenomena. Nowadays, the fundamental challenge of numerical simulation is in designing physical systems while saving the experimentation steps.

As an example, at the early stage of conception in aeronautics, numerical simulation aims at exploring the design parameters space and setting the global variables such that target performances are satisfied. This iterative procedure needs fast multiphysical models. These simplified models are usually calibrated using high-fidelity models or experiments. At each of these levels, modeling requires control of uncertainties due to simplifications of models, numerical errors, data imprecisions, variability of surrounding conditions, etc.

One dilemma in the design by numerical simulation is that many crucial choices are made very early, and thus when uncertainties are maximum, and that these choices have a fundamental impact on the final performances.

Classically, coping with this variability is achieved through *model registration* by experimenting and adding fixed *margins* to the model response. In view of technical and economical performance, it appears judicious to replace these fixed margins by a rigorous analysis and control of risk. This may be achieved through a probabilistic approach to uncertainties, that provides decision criteria adapted to the management of unpredictability inherent to design issues.

From the particular case of aircraft design emerge several general aspects of management of uncertainties in simulation. Probabilistic decision criteria, that translate decision making into mathematical/probabilistic terms, require the following three steps to be considered [58]:

1. build a probabilistic description of the fluctuations of the model's parameters (*Quantification* of uncertainty sources),
2. deduce the implication of these distribution laws on the model's response (*Propagation* of uncertainties),
3. and determine the specific influence of each uncertainty source on the model's response variability (*Sensitivity Analysis*).

The previous analysis now constitutes the framework of a general study of uncertainties. It is used in industrial contexts where uncertainties can be represented by *random variables* (unknown temperature of an external surface, physical quantities of a given material, ... at a given *fixed time*). However, in order for the numerical models to describe with high fidelity a phenomenon, the relevant uncertainties must generally depend on time or space variables. Consequently, one has to tackle the following issues:

- *How to capture the distribution law of time (or space) dependent parameters, without directly accessible data?* The distribution of probability of the continuous time (or space) uncertainty sources must describe the links between variations at neighbor times (or points). The local and global regularity are important parameters of these laws, since it describes how the fluctuations at some time (or point) induce fluctuations at close times (or points). The continuous equations representing the studied phenomena should help *to propose models for the law of the random fields*. Let us notice that interactions between various levels of modeling might also be used to derive distributions of probability at the lowest one.

- The navigation between the various natures of models needs a kind of *metric* which could *mathematically describe the notion of granularity or fineness* of the models. Of course, the local regularity will not be totally absent of this mathematical definition.
- All the various levels of conception, preliminary design or high-fidelity modelling, require *registrations by experimentation* to reduce model errors. This *calibration* issue has been present in this frame since a long time, especially in a deterministic optimization context. The random modeling of uncertainty requires the definition of a systematic approach. The difficulty in this specific context is: statistical estimation with few data and estimation of a function with continuous variables using only discrete setting of values.

Moreover, a multi-physical context must be added to these questions. The complex system design is most often located at the interface between several disciplines. In that case, modeling relies on a coupling between several models for the various phenomena and design becomes a *multidisciplinary optimization* problem. In this uncertainty context, the real challenge turns robust optimization to manage technical and economical risks (risk for non-satisfaction of technical specifications, cost control).

We participate in the uncertainties community through several collaborative research projects (ANR and Pôle SYSTEM@TIC), and also through our involvement in the MASCOT-NUM research group (GDR of CNRS). In addition, we are considering probabilistic models as phenomenological models to cope with uncertainties in the DIGITEO ANIFRAC project. As explained above, we focus on essentially irregular phenomena, for which irregularity is a relevant quantity to capture the variability (e.g. certain biomedical signals, terrain modeling, financial data, etc.). These will be modeled through stochastic processes with prescribed regularity.

4.2. Design of complex systems

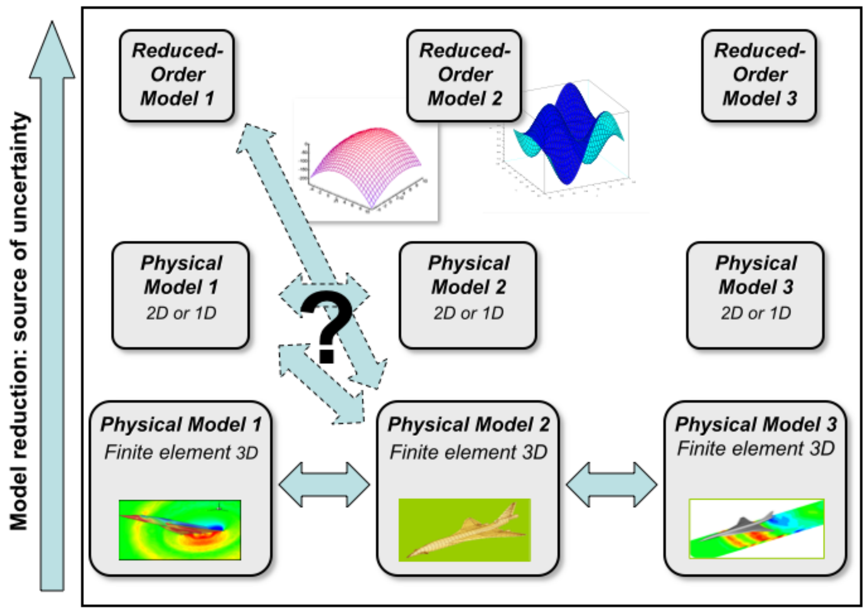

Figure 3. Coupling uncertainty between heterogeneous models

The design of a complex (mechanical) system such as aircraft, automobile or nuclear plant involves numerical simulation of several interacting physical phenomena: CFD and structural dynamics, thermal evolution of a fluid circulation, ... For instance, they can represent the resolution of coupled partial differential equations using finite element method. In the framework of uncertainty treatment, the studied “phenomenological model" is a chaining of different models representing the various involved physical phenomena. As an example, the pressure field on an aircraft wing is the result of both aerodynamic and structural mechanical phenomena. Let us consider the particular case of two models of partial differential equations coupled by limit conditions. The direct propagation of uncertainties is impossible since it requires an exploration and then, many calls to costly models. As a solution, engineers use to build reduced-order models: the complex high-fidelity model is substituted with a CPU less costly model. The uncertainty propagation is then realized through the simplified model, taking into account the approximation error (see [52]).

Interactions between the various models are usually explicited at the finest level (cf. Fig. 3). How may this coupling be formulated when the fine structures of exchange have disappeared during model reduction? How can be expressed the interactions between models at different levels (in a multi-level modeling)? The ultimate question would be: how to choose the right level of modeling with respect to performance requirements?

In the multi-physical numerical simulation, two kinds of uncertainties then coexist: the uncertainty due to substitution of high-fidelity models with approximated reduced-order models, and the uncertainty due to the new coupling structure between reduced-order models.

According to the previous discussion, the uncertainty treatment in a multi-physical and multi-level modeling implies a large range of issues, for instance numerical resolutions of PDE (which do not enter into the research topics of *Regularity*). Our goal is to contribute to the theoretical arsenal that allows to fly among the different levels of modeling (and then, among the existing numerical simulations). We will focus on the following three axes:

- In the case of a phenomenon represented by two coupled partial differential equations whose resolution is represented by reduced-order models, how to define a probabilistic model of the coupling errors? In connection with our theoretical development, we plan to characterize the regularity of this error in order to quantify its distribution. This research axis is supported by an ANR grant (OPUS project).
- The multi-level modeling assumes the ability to choose the right level of details for the models in adequacy to the goals of the study. In order to do that, a rigorous mathematical definition of the notion of *model fineness/granularity* would be very helpful. Again, a precise analysis of the fine regularity of stochastic models is expected to give elements toward a precise definition of granularity. This research axis is supported by a a Pôle SYSTEM@TIC grant (EHPOC project), and also by a collaboration with EADS.
- Some fine characteristics of the phenomenological model may be used to define the probabilistic behaviour of its variability. The action of modeling a phenomena can be seen as an interpolation issue between given observations. This interpolation can be driven by physical evolution equations or fine analytical description of the physical quantities. We are convinced that Hölder regularity is an essential parameter in that context, since it captures how variations at a given point induce variations at its neighbors. Stochastic processes with prescribed regularity (see section 3.3) have already been used to represent various fluctuating phenomena: Internet traffic, financial data, ocean floor. We believe that these models should be relevant to describe solutions of PDE perturbed by uncertain (random) coefficients or limit conditions. This research axis is supported by a Pôle SYSTEM@TIC grant (CSDL project).

The preliminary design of complex systems can be described as an exploration process of a so-called design space, generated by the global parameters. An interactive exploration, with a decisional visualization goal, needs reduced-order models of the involved physical phenomena. We are convinced that the local regularity of phenomena is a relevant quantity to drive these approximated models. Roughly speaking, in order to be representative, a model needs more informations where the fluctuations are the more important (and consequently, where irregularity is the more important).

In collaboration with Dassault Aviation, EDF and EADS, we study how the local regularity can provide a good quantification of the concept of *granularity* of a model, in order to select the good level of fidelity adapted to the requiered precision.

Our works in that field can be expressed into:

- The definition and the study of stochastic partial differential equations driven by processes with prescribed regularity (that do not enter into the classical theory of stochastic integration).
- The study of the evolution of the local regularity inside stochastic partial differential equations (SPDE). The stochastic 2-microlocal analysis should provide informations about the local regularity of the solutions, in function of the coefficients of the equations. The knowledge of the fine behaviour of the solution of the SPDE will provide important informations in the view of numerical simulations.

4.3. Biomedical Applications

ECG analysis and modelling

ECG and signals derived from them are an important source of information in the detection of various pathologies, including *e.g.* congestive heart failure, arrhythmia and sleep apnea. The fact that the irregularity of ECG bears some information on the condition of the heart is well documented (see *e.g.* the web resource http://www.physionet.org). The regularity parameters that have been studied so far are mainly the box and regularization dimensions, the local Hölder exponent and the multifractal spectrum [61], [63]. These have been found to correlate well with certain pathologies in some situations. From a general point of view, we participate in this research area in two ways.

- First, we use refined regularity characterizations, such as the regularization dimension, 2-microlocal analysis and advanced multifractal spectra for a more precise analysis of ECG data. This requires in particular to test current estimation procedures and to develop new ones.
- Second, we build stochastic processes that mimic in a faithful way some features of the dynamics of ECG. For instance, the local regularity of RR intervals, estimated in a parametric way based on a modelling by an mBm, displays correlations with the amplitude of the signal, a feature that seems to have remained unobserved so far [3]. In other words, RR intervals behave as SRP. We believe that modeling in a simplified way some aspects of the interplay between the sympathetic and parasympathetic systems might lead to an SRP, and to explain both this self-regulating property and the reasons behind the observed multifractality of records. This will open the way to understanding how these properties evolve under abnormal behaviour.

Pharmacodynamics and patient drug compliance

Poor adherence to treatment is a worldwide problem that threatens efficacy of therapy, particularly in the case of chronic diseases. Compliance to pharmacotherapy can range from 5% to 90%. This fact renders clinical tested therapies less effective in ambulatory settings. Increasing the effectiveness of adherence interventions has been placed by the World Health Organization at the top list of the most urgent needs for the health system. A large number of studies have appeared on this new topic in recent years [77], [76]. In collaboration with the pharmacy faculty of Montréal university, we consider the problem of compliance within the context of multiple dosing. Analysis of multiple dosing drug concentrations, with common deterministic models, is usually based on patient full compliance assumption, *i.e.*, drugs are administered at a fixed dosage. However, the drug concentration-time curve is often influenced by the random drug input generated by patient poor adherence behaviour, inducing erratic therapeutic outcomes. Following work already started in Montréal [70], [71], we consider stochastic processes induced by taking into account the random drug intake induced by various compliance patterns. Such studies have been made possible by technological progress, such as the "medication event monitoring system", which allows to obtain data describing the behaviour of patients.

We use different approaches to study this problem: statistical methods where enough data are available, model-based ones in presence of qualitative description of the patient behaviour. In this latter case, piecewise deterministic Markov processes (PDP) seem a promising path. PDP are non-diffusion processes whose evolution follows a deterministic trajectory governed by a flow between random time instants, where it undergoes a jump according to some probability measure [56]. There is a well-developed theory for PDP, which studies stochastic properties such as extended generator, Dynkin formula, long time behaviour. It is easy to cast a simplified model of non-compliance in terms of PDP. This has allowed us already to obtain certain properties of interest of the random concentration of drug [44]. In the simplest case of a Poisson distribution, we have obtained rather precise results that also point to a surprising connection with infinite Bernouilli convolutions [44], [11], [10]. Statistical aspects remain to be investigated in the general case.

SECSI Project-Team

4. Application Domains

4.1. Application Domains

Here are a few examples of applications of research done in SECSI:

- Security of electronic voting schemes: the case of the Helios protocol, used in particular at University of Louvain-la-Neuve (2010) and at the International Association for Cryptographic Research (IACR).
- Security of the protocols involved in the TPM (Trusted Platform Module) chip, a chip present in most PC laptops today, and which is meant to act as a trusted base.
- Security of the European electronic passport—and the discovery of an attack on the French implementation of it.
- The Tookan tool allows one to assess the security of security tokens. These tokens are meant as safes holding secret keys, which should never be permitted to get out unencrypted. Several vulnerabilities discovered. Several interesting customers in banking (Barclays), in aeronautics (Boeing), notably.
- Intrusion detection with the Orchids tool: several interested partners, among which EADS Cassidian, Thales, Galois Inc. (USA), the French Direction Générale de l'Armement (DGA).

SELECT Project-Team

4. Application Domains

4.1. Introduction

A key goal of SELECT is to produce methodological contributions in statistics. For this reason, the SELECT team works with applications that serve as an important source of interesting practical problems and require innovative methodologies to address them. Most of our applications involve contracts with industrial partners, e.g. in reliability, although we also have several more academic collaborations, e.g. genomics, genetics, neuroimaging and ancient material imaging.

4.2. Curves classification

The field of classification for complex data as curves, functions, spectra and time series is important. Standard data analysis questions are being revisited to define new strategies that take the functional nature of the data into account. Functional data analysis addresses a variety of applied problems, including longitudinal studies, analysis of fMRI data and spectral calibration.

We are focusing on unsupervised classification. In addition to standard questions as the choice of the number of clusters, the norm for measuring the distance between two observations, and the vectors for representing clusters, we must also address a major computational problem. The functional nature of the data needs to be design efficient anytime algorithms.

4.3. Computer Experiments and Reliability

Since several years, SELECT has collaborations with EDF-DER *Maintenance des Risques Industriels* group. An important theme concerns the resolution of inverse problems using simulation tools to analyze incertainty in highly complex physical systems. A collaboration on an analogous topic is developed with Dassault Aviation.

The other major theme concerns probabilistic modeling in fatigue analysis in the context of a research collaboration with SAFRAN an high-technology group (Aerospace propulsion, Aicraft equipment, Defense Security, Communications).

4.4. Neuroimaging

Since 2007 SELECT participates to a working group with team Neurospin (CEA-INSERM-Inria) on Classification, Statistics and fMRI (functional Magnetic Resonance Imaging) analysis. In this framework two theses have been co-supervised by SELECT and Neurospin researchers (Merlin Keller 2006-2009 and Vincent Michel 2007-2010). The aim of this research is to determine which parts of the brain are activated by different types of stimuli. A model selection approach is useful to avoid "false-positive" detections.

4.5. Analysis of genomic data

For the past few years SELECT has collaborated with Marie-Laure Martin-Magniette (URGV) for the analysis of genomic data. An important theme of this collaboration is using statistically sound model-based clustering methods to discover groups of co-expressed genes from microarray and high-throughput sequencing data. In particular, identifying biological entities that share similar profiles across several treatment conditions, such as co-expressed genes, may help identify groups of genes that are involved in the same biological processes.

4.6. Environment

A study has been achieved by Jean-Michel Poggi, Michel Misiti, Yves Misiti and Bruno Portier (INSA de Rouen), in the context of a collaboration between AirNormand, Orsay University and INSA of Rouen. Mixtures of linear regression models are used for the short-term statistical forecasting of the daily mean PM10 concentration in three cities in Haute-Normandie (France): Rouen, Le Havre and Dieppe. The Haute-Normandie region is located at northwest of Paris, near the south side of Manche sea and is heavily industrialized. Six monitoring stations reflecting the diversity of situations: urban background, traffic, rural and industrial stations are considered. This recent statistical method has been used and beyond the application, this study shed light on this method [35].

4.7. Analysis spectroscopic imaging of ancient materials

Ancient materials, encountered in archaeology, paleontology and cultural heritage, are often complex, heterogeneous and poorly characterised before their physico-chemical analysis. A technique of choice to gather as much physico-chemical information as possible is spectro-microscopy or spectral imaging where a full spectra, made of more than thousand samples, is measured for each pixel. The produced data is tensorial with two or three spatial dimensions and one or more spectral dimensions and it requires the combination of an «image» approach with «curve analysis» approach. Since 2010 SELECT collaborates with Serge Cohen (IPANEMA) on the development of conditional density estimation through GMM and non-asymptotic model selection to perform stochastic segmentation of such tensorial dataset. This technic enablesx the simultaneous accounting for spatial and spectral information while producing statistically sound information on morphological and physico-chemical aspects of the studied samples.

TAO Project-Team

4. Application Domains

4.1. Energy Management

Energy management, our prioritary application field, involves sequential decision making with:

- stochastic uncertainties (typically weather);
- both high scale combinatorial problems (as induced by nuclear power plants) and non-linear effects;
- high dimension (including hundreds of hydroelectric stocks);
- multiple time scales:
  - minutes (dispatching, ensuring the stability of the grid), essentially beyond the scope of our work, but introducing constraints for our time scales;
  - days (unit commitment, taking care of compromises between various power plants);
  - years, for evaluating marginal costs of long term stocks (typically hydroelectric stocks);
  - tenths of years, for investments.

Nice challenges also include:

- spatial distribution of problems; due to capacity limits we can not consider a power grid like Europe + North Africa as a single “production = demand” constraint; with extra connections we can equilibrate excess production by renewables for remote areas, but no in an unlimited manner.
- other uncertainties, which might be modelized by adversarial or stochastic frameworks (e.g. technological breakthroughs, decisions about ecological penalization).

We have several related projects (Citines, a European (FP7) project; in the near future we should start the Post project (Ademe); IOMCA, a ANR project). We have a collaboration with a company, Artelys, working on optimization in general, and in particular on energy management.

Technical challenges: Our work focuses on the combination of reinforcement learning tools, with their anytime behavior and asymptotic guarantees, with existing fast approximate algorithms; see 6.4 . Our goal is to extend the state of the art by taking into account non-linearities which are often neglected in power systems due to the huge computational cost.

Related Activities:

- We are in the process of creating a Franco-Taiwanese company (maybe a taiwanese company using French software) for energy optimization in Taiwan.
- We have a joint team with Taiwan, namely the Indema associate team (see Section 8.4.1.1).
- We have a “Ilab” in progress with Artelys (see Section 5.1) for industrialization of our work.
- We organized various forums and meetings around Energy Management.

4.2. Air Traffic Control

Air Traffic Control has been an application field of Marc Schoenauer’s work in the late 90s (PhD theses of F. Médioni in 98 and S. Oussedik in 2000). It was revived recently with Gaëtan Marceau-Caron’s CIFRE PhD together with Thalès Air Systems (Areskin HAdjaz) and Thalès TRT (Pierre Savéant), around global optimization of the traffic in order to increase the capacity of the airspace without overloading the controllers. First results concern the on-line prediction of individual plane trajectories [42]: these results demonstrate that there is no hope to ever predict the global behavior of the traffic by considering planes individually. A new formulation of the problem is hence considered, where plane flows are modeled with Bayesian Networks [42]. On-going work is concerned with the simulation and optimization of air traffic based on this model.

TOCCATA Team

4. Application Domains

4.1. Application Domains

Keywords: embedded software, smartcards, avionics, telecommunication, transportation systems

The application domains we target involve safety-critical software, that is where a high level guarantee of soundness of functional execution of the software is wanted. The domains of application include

- Transportation: aeronautics, railroad, space flight, automotive
- Communications: mobile phones, smart phones, Web applications
- Financial applications, banking
- Medicine: diagnostic devices, computer-assisted surgery
- Databases with confidentiality requirements (e.g. health records, electronic voting)

Currently our industrial collaborations mainly belong the first of these domains: transportation. These include, in the context of the ANR U3CAT project (Airbus France,Toulouse; Dassault Aviation, Saint-Cloud; Sagem Défense et Sécurité):

- proof of C programs via *Frama-C*/*Jessie*/*Why* ;
- proof of floating-point programs ;
- use of the *Alt-Ergo* prover via CAVEAT tool (CEA) or *Frama-C*/WP.

In the context of the FUI project Hi-Lite, the Adacore (Paris) uses *Why3* and *Alt-Ergo* as back-endd to GnatProve, an environment for verification of Ada programs. This is applied in the domain of aerospace (Thales).

In the context of a new ANR project BWare, we investigate the use of *Why3* and *Alt-Ergo* as an alternative back-end for checking proof obligation generated by *Atelier B*, whose main applications are railroad-related software (http://www.methode-b.com/documentation_b/ClearSy-Industrial_Use_of_B.pdf, collaboration with Mitsubishi Electric R&D Centre Europe, Rennes; ClearSy, Aix-en-Provence)

Apart from the domain of transportation, the Cubicle model checker modulo theoriesbased on the *Alt-Ergo* SMT prover (collaboration with Intel Strategic Cad Labs, Hillsboro, OR, USA) can be applied to verification of concurrent programs and protocols (http://cubicle.lri.fr/).

TYPICAL Project-Team (section vide)