RESEARCH CENTER
Saclay - Île-de-France

FIELD

Activity Report 2013

Section Application Domains

Edition: 2014-03-19

AMIB Project-Team (section vide)

AVIZ Project-Team

4. Application Domains

4.1. Panorama

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);

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Nicolas Bordenabe, Konstantinos Chatzikokolakis, Catuscia Palamidessi, Marco Stronati.

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 (section vide)

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 in collaboration with SAGEM.

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) in collaboration with St Antoine Hospital in Paris.

4.3. Energy Management

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

GALEN Project-Team

4. Application Domains

4.1. Brain Tumors and Neuro-degenerative diseases

: 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. Furthermore, in collaboration with the Neurospin center of CEA and the Brookhaven National Laboratory at StonyBrook University we investigate the use of machine learning methods towards automatic interpretation of functional magnetic resonance imaging between cocaine addicted and normal subjects. Last, but not least in collaboration with the Georges Pompidou European Hospital an effort toward understanding tumor perfusion process through comportemental models is carried out with emphasis given on elastic organs.

4.2. Image-driven Radiotherapy Treatment & Surgery Guidance

The use of CT and MR imaging for cancer guidance treatment in collaboration with the Gustave Roussy Institute of Oncology. 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 prior to each session and the ones used for scheduling/planning and dose estimation. 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.

GECO Project-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, [89], [94]). The rapid evolution of the domain is driven by a multitude of experiments getting more and more precise and complex (see the recent review [49]). 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 [31],
- bounds on the controllability time [27],
- STIRAP processes [99],
- simultaneous control [72],
- optimal control ([68], [40], [51]),
- numerical simulations [78].

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, [32], [69], [90], [46]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [36], [47], [65], [37]). 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 [98]. Moreover, it has been shown that a relevant model, the quantum oscillator, is not even approximately controllable [91], [81]. These negative results have been more recently completed by positive ones. In [38], [39] 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 [54]). 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 [80], [85], [66]. While [80] studies a controlled Schrödinger equation in $\mathbb{R}$ for which the uncontrolled Schrödinger operator has mixed spectrum, [85], [66] 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 [57], [43].

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 [63]. Optimal control approaches have also been considered [35], [48]. 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 [87]) 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 ([88]) and then modified by Citti and Sarti ([53]). The model is based on experimental observations, and in particular on the fundamental work by Hubel and Wiesel [62] 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 robotics and neurophysiology. It could help to design better control strategies for robots and artificial limbs, yielding smoother and more progressive movements. Another 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 [34]), 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 [97] 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 [67] or for Markov processes [86]. 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 ([74]).

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 [95], [75]).

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 [76]. 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 [82] and, provided that the admissible modes are finitely many, the Lyapunov function can be taken polyhedral or polynomial [41], [42], [55]. 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 [56] and references therein). Algebraic approaches to prove the stability of switched systems under arbitrary switching, not relying on Lyapunov techniques, have been proposed in [73], [28].

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 [61].

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 [70]. It is known that, in this context, stability cannot be tested on periodic signals alone [44].

Finally, let us mention that little is known about infinite-dimensional switched system, with the exception of some results on uniform asymptotic stability ([79], [92], [93]) and some recent papers on optimal control ([60], [100]).

GEOMETRICA Project-Team

4. Application Domains

4.1. Application Domains

- Medical Imaging
- Numerical simulation
- Geometric modeling
- Geographic information systems
- Visualization
- Data analysis
- Astrophysics
- Material physics

GRACE Project-Team

4. Application Domains

4.1. Cryptography and Cryptanalysis

In the twenty-first century, cryptography plays two essential rôles: it is used to ensure *security* and *integrity* of communications and communicating entities. Contemporary cryptographic techniques can be used to hide private data, and to prove that public data has not been modified; to provide anonymity, and to assert and prove public identities. The creation and testing of practical cryptosystems involves

1. The design of provably secure protocols;
2. The design and analysis of compact and efficient algorithms to implement those protocols, and to attack their underlying mathematical and computational problems; and
3. The robust implementation of those algorithms in low-level software and hardware, and their deployment in the wild.

While these layers are interdependent, GRACE's cryptographic research is focused heavily on the middle layer: we design, implement, and analyze the most efficient algorithms for fundamental tasks in contemporary cryptography. Our "clients", in a sense, are protocol designers on the one hand, and software and hardware engineers on the other.

François Morain and Benjamin Smith work primarily on the number-theoretic algorithms that underpin the current state-of-the-art in public-key cryptography (which is used to establish secure connections, and create and verify digital signatures, among other applications). For example, their participation in the ANR CATREL project aims to give a realistic assessment of the security of systems based on the Discrete Logarithm Problem, by creating a free, open, algorithmic package implementing the fastest known algorithms for attacking DLP instances. This will have an extremely important impact on contemporary pairing-based cryptosystems, as well as legacy finite field-based cryptosystems. On a more constructive note, Morain's elliptic curve point counting and primality proving algorithms are essential tools in the everyday construction of strong public-key cryptosystems, while Smith's recent work on elliptic curves aims to improve the speed of curve-based cryptosystems (such as Elliptic Curve Diffie–Hellman key exchange, a crucial step in establishing secure internet connections) without compromising their security.

Daniel Augot, Françoise Levy-dit-Vehel, and Alain Couvreur's research on codes has far-reaching applications in *code-based cryptography*. This is a field which is growing rapidly in importance—partly due to the supposed resistance of code-based cryptosystems to attacks from quantum computing, partly due to the range of new techniques on offer, and partly because the fundamental problem of parameter selection is relatively poorly understood. For example, Couvreur's work on distinguishing codes has an important impact on the design of code-based systems built over algebraic geometry codes, and on the choice of parameter sizes for secure implementations. But coding theory also has important practical applications in the improvement of conventional symmetric cryptosystems. For example, Augot's recent work on MDS matrices via BCH codes gives a more efficient construction of optimal diffusion layers in block ciphers.

4.2. Privacy

While cryptography classically aims to provide confidentiality for messages during their transmission between a sender and a recipient, privacy is a broader, more subtle, and sometimes less technical issue.

Daniel Augot with other groups from Inria (Comete, SMIS) started discussions with lawyers and economists, fostered by IDEX Paris-Saclay's *Institut de la société du numérique*, to understand the privacy concerns of ordinary citizens. On a more technical side, privacy can be protected with cryptographic protocols other than encryption. In this direction, Grace is engaged since April 2013 in a collaboration with Alcatel–Lucent on private data storage and retrieval in the cloud.

GRAND-LARGE Project-Team (section vide)

IN-SITU Project-Team

4. Application Domains

4.1. Application 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.

M3DISIM Team

4. Application Domains

4.1. Clinical applications

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [8]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [1].

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 [77], des problèmes de calcul de débit pour des ateliers [126], pour des réseaux ferroviaires [76] ou routiers, et l'évaluation de performance des réseaux de communication [66]. 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" [70], [135]. Le projet Maxplus collabore avec le projet Metalau, qui étudie particulièrement les applications des modèles max-plus à la modélisation microscopique du trafic routier [143], [140], [102].

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 [77], 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 [76] and [66]. 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 [70], [135]. The Maxplus team collaborates with the Metalau team, working particularly on the applications of max-plus models to the microscopic modelling of road traffic [143], [140], [102].

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 [78]. Ces liens conduisent parfois à de nouveaux algorithmes pour les problèmes de recherche opérationnelle classiques,

comme le problème de circuit de poids moyen maximum [85]. Certains problèmes combinatoires, comme des problèmes de programmation disjonctive, peuvent être décomposés par des méthodes de type max-plus [176]. 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 [112] pour le cas du "jobshop", une représentation plus simple a été obtenue dans [137] 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 [137]. Finalement, en regardant l'algèbre max-plus comme une limite de l'algèbre classique, on peut utiliser des outils algébriques en optimisation combinatoire [133].

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 [78]. Sometimes, these relations lead to new algorithms for classical Operations Research problems, like the maximal circuit mean [85]. There are also special combinatorial problems, like certain problems of disjunctive programming, which can be decomposed by max-plus type methods [176]. 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 [112] for the jobshop case, a simpler representation was given in [137] 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 [137]. Finally, viewing max-plus algebra as a limit of classical algebra allows to use algebraic tools in combinatorial optimisation [133].

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

L'interprétation abstraite est une technique, introduite par P. et R. Cousot [89], 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 [88], [109], 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 [90], 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 [110], [83], [84],[7]. Une version assez générale de cet algorithme, adaptée au domaine des templates, est décrite dans [109] 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 [89], 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 [88], [109], 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 [90], 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 [110], [83], [84],[7]. A rather general version of this policy iteration algorithm, adpated to the domain of templates, is described in [109], 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 [87].

English version

Max-plus algebra arises naturally in the computation of similarity scores, in biological sequence comparison. See for instance [87].

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

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.

We have participated in the Univerself Project (see below) on self-aware networks, and will be searching new cooperations.

4.2. Transport Systems

Participants: Stefan Haar, Serge Haddad, Simon Theissing.

We participate in the project MIC on multi-modal transport systems with in the IRT *System X*, with academic partners UPMC, IFSTTAR and CEA, and several industrial partners including Alstom (project leader), COSMO and Renault. Transportation operators in an urban area need to plan, supervise and steer different means of transportation with respect to several criteria:

- Maximize capacity;
- guarantee punctuality and robustness of service;
- minimize energy consumption.

The systems must achieve these objectives not only under ideal conditions, but also be robust to perturbations (such as a major cultural or sport event creating additional traffic), modifications of routes (roadwork, accidents, demonstrations, ...) and tolerant to technical failures. Therefore, systems must be enabled to raise appropriate alarms upon detection of anomalies, diagnose the type of anomaly and select the appropriate response.

While the above challenges belong already to the tasks of individual operators in the unimodal setting, the rise of and increasing demand for *multi-modal* transports forces to achieve these planning, optimization and control goals not in isolation, but in a cooperative manner, across several operators. The research task here is first to analyze the transportation system regarding the available means, capacities and structures, and so as to identify the impacting factors and interdependencies of the system variables. Based on this analysis, the task is to derive and implement robust planning, with tolerance to technical faults; diagnosis and control strategies that are optimal under several, possibly different, criteria (average case vs worst case performance, energy efficiency, etc.) and allow to adapt to changes e.g. from nominal mode to reduced mode, sensor failures, etc.

OAK Project-Team

4. Application Domains

4.1. Business Intelligence for Open Data

Research developed in the group helps publish, curate, and exploit open data, in particular the data produced by local or national administrations and which is returned to the general public under the form of applications (often Web-based, often mobile) which increase opportunities for business or leisure. This concerns in particular our work on Open Data entity resolution [36] and Open Data warehousing [33], [32]. This research is set to be deployed on real Open Data sets from the Grenoble urban area, within the industry-led Datalyse project (Section 7.2.1).

4.2. Social Data Management

We develop models and algorithms for efficiently exploiting, enhancing, and querying social network data, in particular based on structured content, semantic annotations, and user interaction networks. We pursue this research with many industrial partners within the ALICIA project (Section 7.2.1) as well as in the Structured, Social, and Semantic Search project (Section 7.2.2).

4.3. Data Journalism

Efficiently handling the deluge of news and other news-worthy electronic data being published today, requires powerful content management tools in order to handle news document structure, extract meaning from the text, connect pieces of information with each other, etc. To that purpose, we have built and experimented with FactMinder, a platform for gathering, enriching, annotating, storing, and querying news documents, with the help of existing ontologies that users may enrich and/or exploit next to their own [24]. Many more applications of our research are possible in this domain [38].

4.4. Data Transformation Debugging

All applications mentioned above, e.g., business intelligence, data integration, or data enrichment in social data management or data journalism take as input some data to be further manipulated and transformed. In many applications, including again business intelligence and data journalism, the correctness of the produced output data is crucial. It is thus important to verify the semantic correctness of a data transformation and to be able to trace back what has happened to the data within the transformation. We support this data transformation debugging based on provenance [31], [26].

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 $\mathbf{X}$ 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, $\mathbf{Y}$ 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 $\mathbf{Y}_1, .., \mathbf{Y}_{n_f}$ of Y independently, yielding n_f problems to be solved in parallel:

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

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

$$\widehat{\beta_i} = \text{argmin}_{\beta_i} \|\mathbf{Y}_i - \mathbf{X}\beta_i\|^2 + \Psi(\beta_i), \tag{2}$$

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

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

with λ_1, λ_2, η_1, $\eta_2 \geq 0$. In general, only one or two of these constraints is considered (hence is enforced with a non-zero coefficient):

- 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 $\mathbf{Y}_i$ 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.

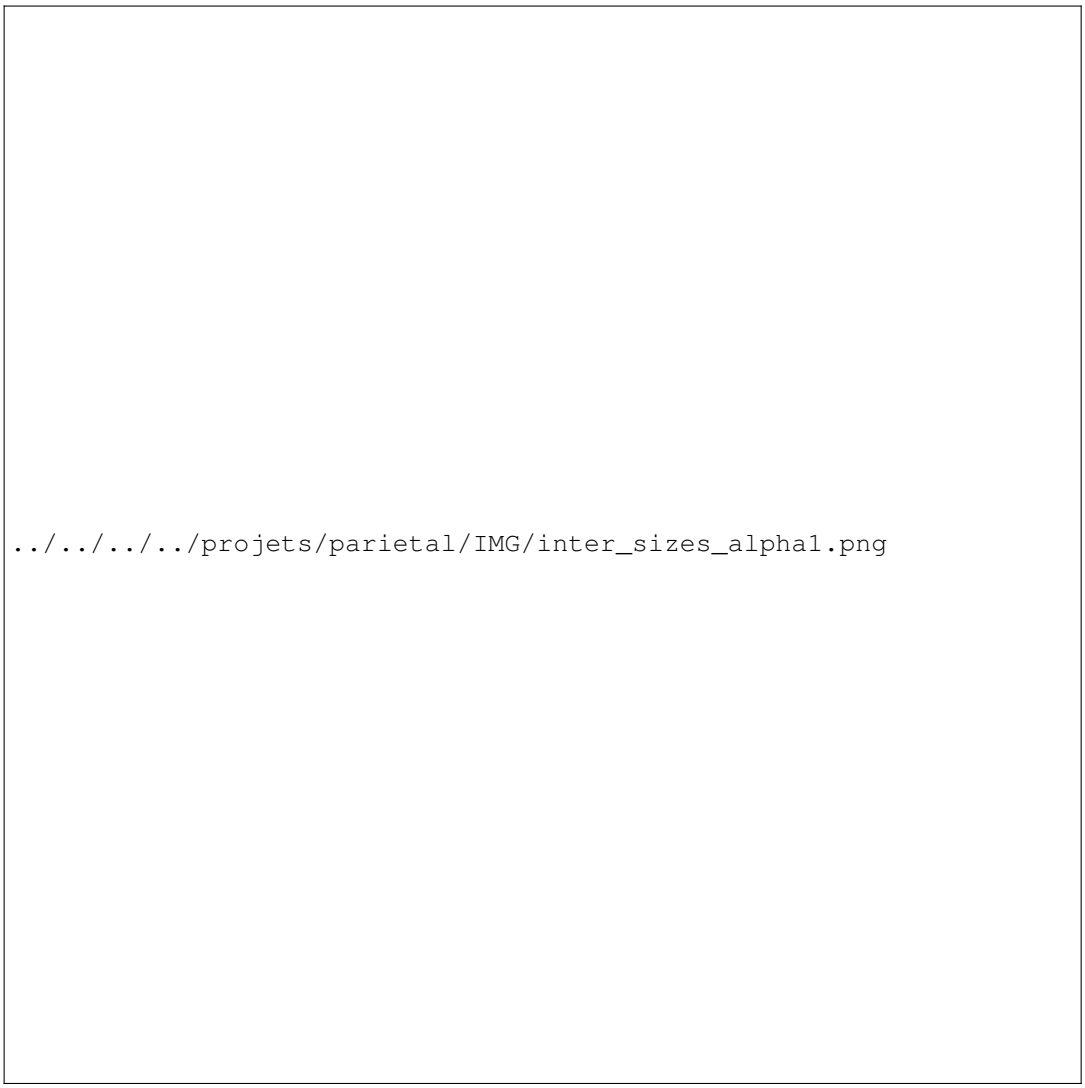

Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problems here consists in predicting the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the observation of an image. Learning and test are performed 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 logistic 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 $\mathbf{Y}_i$ is not well fitted by the model, the estimation of other variables $\mathbf{Y}_j, j \neq i$ may provide constraints on the support of β_i and thus, improve the prediction of $\mathbf{Y}_i$. Yet this does not impose constraints on the non-zero parameters of the parameters β_i.

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

then

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

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

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

where $\mathbf{D}$ represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}), and $\mathbf{A}$ the associated subject-wise loadings. While traditional PCA and independent components analysis are limited to reconstruct components $\mathbf{D}$ within the space spanned by the column of $\mathbf{X}$, it seems desirable to add some constraints on the rows of $\mathbf{D}$, 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:

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

where $(\mathbf{A}_i),\ i \in \{1..n_f\}$ represents the columns of $\mathbf{A}$. 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 $\mathbf{D}$ for $\mathbf{A}$ fixed, and conversely. This readily suggests an alternate optimization scheme, where $\mathbf{D}$ and $\mathbf{A}$ 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 consider different means of parametrizing covariance distributions and how these parametrizations impact the test of statistical differences across individuals. 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 Reasoning

Automated reasoning has traditionally focused on classical first-order logic but it is increasingly important for automation to other logics. We are applying our research to the following extensions to this traditional focus.

- 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 [38] continue to be the fastest automated theorem provers for propositional and first-order linear logic.
- 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 focused logic. One of the advantages of this approach is that it allows combining and extending such techniques. For example, the PSI project has applied this approach to proof to the problem of SAT-modulo-Theory. In that domain, logical reasoning is combined with domain-specific decision procedures. The PSI project has shown how to incorporate the call to decision procedures in the proof-theoretical framework of focused sequent calculi and the proof-search mechanisms that are related to it.

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. There is also the POPLMark challenge [37] which 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 been applying their research results to design and building systems to directly aid in both of these challenges. One important requirements for reasoning about 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. Also, 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

Within the context of the ProofCert project, various members of the team have been building a flexible framework for the definition of the semantics of proof evidence. The emphasis is to attempt to capture as many forms of proof evidence as is possible. Using this framework, we have defined the semantics of all the following forms of proof evidence: natural deduction, expansion trees, matings, proof nets, resolution refutations, and Frege proofs. Given our framework, there is one kernel that can check all of these different forms of proof. Thus, one only needs to trust this one kernel in order to trust the output of a very wide range of theorem provers working in either intuitionistic or classical logics (see [20], [19], and [32].

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 modelling (Euler equations, Galbrun equations, Goldstein equation) 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 volumic 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 modelling and simulation of the timpani, the guitar and the piano have been carried out in collaboration with A. Chaigne of ENSTA. This work will continue in the framework of the European Project BATWOMAN.

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 microcomponents 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 Team

4. Application Domains

4.1. Pharmacometrics

Participants: Marc Lavielle, Kevin Bleakley, Célia Barthélémy, Hector Mesa, Elodie Maillot, 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
- protocol optimization tools
- diagnostic tools
- model selection tools
- data exploration tools
- estimation techniques for complex models (eg, stochastic differential equations, partial differential equations)

4.2. Pharmacogenetics

Participants: Marc Lavielle, Kevin Bleakley, Célia Barthélémy.

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. POPIX is working to develop methods for simultaneous model selection and parameter estimation in the SAEM framework in such cases.

4.3. Oncology

Participants: Marc Lavielle, Célia Barthélémy.

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 project Lab 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 ongoing, with the aim of extending the statistical methods developed by POPIX to partial differential equation-based models. NUMED works on models of tumor growth and has previously implemented an extension of MONOLIX to KPP-type reaction-diffusion models.

4.4. Respiratory system

Participants: Bertrand Maury, Astrid Decoene.

Comprehensive models to simulate the whole pulmonary system, i.e., the mechanical behavior of the lung and gas exchanges within the pulmonary system, are built upon ODE and PDE approaches. For instance, the mechanical behavior of a lung is often described by single or multi-compartment ODE models, whereas air flow may be determined by the coupling of a 3D PDE system in the proximal part of the bronchial tree with a 0D ODE system in the distal part of the bronchial tree. Gas exchange has so far been investigated using 0D or 1D models in which heterogeneity of gas exchange along the path length may be investigated.

In a mathematical representation of such physiological systems, model parameters can be associated with specific quantities in the real system, such as the resistance and compliance of the pulmonary system. These quantities are time-dependent and nonlinear and are measured by pneumologists in order to characterize chronic obstructive pulmonary diseases (COPD) such as asthma and emphysema. These parameters may be useful in assessing lung conditions.

Although most physiological studies have used averaged deterministic models of the tracheobronchial tree geometry, morphometric studies show that inter-subject and intra-subject variability in the structural components of the human lung is significant. In particular, the resistance of the respiratory tract may be significantly affected as it is directly related to the inner diameter of the bronchi. Feedback from such variability to resistance and, as a consequence efficiency of the gas exchange process, within the framework of a fully coupled model, is unclear. In this situation, the statistical and numerical approaches being developed by POPIX are clearly promising estimation methods for respiratory system analysis.

4.5. Blood flow modeling

Participants: Bertrand Maury, Astrid Decoene.

Modeling and numerical simulation of blood flow in arteries and veins may become an important tool for medical applications, as for instance in the prediction of cardiovascular disease. Analyzing the pressure waves and estimating the wall compliance of arteries is fundamental, as these exhibit strong inter- and intra-subject variability. Currently, non-invasive pressure measurements involve excessive errors; intensive direct estimation is thus not applicable in practice. Physiologists therefore hope to be able to predict the time and space evolution of the pressure in the arterial network from a small amount of flow data measured at a few points.

Several numerical models have been developed in order to simulate blood flow in arteries and veins. They mainly consist of one to three-dimensional systems of partial differential equations, depending on the level of complexity one desires to achieve. Coupling the various models is also an issue. These numerical models allow us to compute the transversal section area, as well as the velocity or flow at different points in space, leading to a rather complete description of the arterial flow (velocity, pressure, section). But for these models to be adapted to each patient, certain numerical and physical parameters must be fitted, such as the compliance of walls and the viscosity of the blood. These parameters are difficult to estimate experimentally and may be related to measurements which involve a non-negligible error. Furthermore, their optimal value is linked to the particular modeling framework and therefore can differ from the value given by their physical definition.

Mixed models appear to be an appropriate framework for taking into account the specific nature of each patient and quantifying uncertainty in the numerical model. Flow data are available as it is possible to non-invasively measure the mean velocity in and diameter of an artery.

We aim to introduce statistical mixed models to the framework for the classical one-dimensional blood flow model.

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. 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. 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 [75], [74]. 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 [67], [68], 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 [55]. 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 [37]. 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 [37], [13], [12]. 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.
- 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 and neuroimaging.

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.

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).

Moreover, a collaboration has started with Dassault Aviation on modal analysis of mechanical structures, which aims at identifying the vibration behavior of structures under dynamic excitations. From algorithmic view point, modal analysis amounts to estimation in parametric models on the basis of measured excitations and structural responses data. As it appears from literature and existing implementations, the model selection problem attached to this estimation is currently treated by a rather heavy and very heuristic proced ure. The model selection via penalisation tools are intended to be tested on this model selection problem.

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

Since many years SELECT collaborates 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. Yann Vasseur started a thesis cosupervised by Gilles Celeux and Marie-Laure Martin-Magniette on this topic which is also an interesting investigation domain for the latent block model developed by SELECT.

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. Two methods for spatial outlier detection have been considered: one based on the nearest neighbours weighted median and one based on kriging increments instead of more traditional pseudo-innovations. The two methods are applied to the PM10 monitoring network in Normandie (France) and are fully implemented in the Measurements Quality Control process.

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.

Specfun Team

4. Application Domains

4.1. Experimental mathematics with special functions

Applications in combinatorics and mathematical physics frequently involve equations of so high orders and so large sizes, that computing or even storing all their coefficients is impossible on existing computers. Making this tractable is another challenge of our project. The approach we believe in is to design algorithms of good, ideally quasi-optimal, complexity in order to extract precisely the required data from the equations, while avoiding the computationally intractable task of completely expanding them into an explicit representation.

Typical applications with expected high impact are the automatic discovery and proof of results in combinatorics and mathematical physics for which human proofs are currently unattainable.

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 cannot consider a power grid like Europe + North Africa as a single "production = demand" constraint; with extra connections we can balance excess production by renewables for remote areas, although to a limited extent.
- Other uncertainties, which might be modeled 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), and POST, a ADEME project about investments in power systems. We have a collaboration with the SME Artelys, that works on optimization in general, and on energy management in particular.

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 6.2 . 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 an I-lab in progress with Artelys (see Section 5.1) in order to ensure the transfer of our work.
- We have 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 1998 and S. Oussedik in 2000). It was revived recently with Gaëtan Marceau-Caron's CIFRE PhD together with Thalès Air Systems (Areski Hadjaz) and Thalès TRT (Pierre Savéant), tackling the global optimization of the traffic in order to increase the capacity of the airspace without overloading the controllers. A new formulation of the problem, modeling the plane flows with Bayesian Networks, has been proposed to the Air Traffic Control community [48], [50]. The goal of the optimization is to minimize the cumulated delays of all flights, while maintaining a reasonnable level of congestion in all sectors. These objectives are computed using Monte-Carlo simulations of the Bayesian network, and Evolutionary Algorithms are used to address the resulting stochastic multi-objective optimization problem [49].

TOCCATA Team

4. Application Domains

4.1. Mission-Critical Software

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-end to GnatProve, an environment for verification of Ada programs. This is applied in the domain of aerospace (Thales, EADS Astrium).

In the context of 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 theories based 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/).