Activity Report 2015

Section Application Domains

AMIB Project-Team (section vide)

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

4. Application Domains

4.1. Application Domains

Research in visual analytics can profit from the challenges and requirements of real-world datasets. Aviz develops active collaboration with users from a range of 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.

We apply our techniques to important medical applications domains such as bioinformatics and brain studies. In particular, we are interested in helping neuroscientists make sense of evolving functional networks, in the form of weighted and/or dynamic graphs.

Other application domains include:

- Digital Humanities in general, with the Cendari European project with historians from most European countries, the project "Interactive Network Visualization" with Microsoft Research-Inria Joint Centre on Graph Visualization, and with our work on Word-Scale Visualizations;

- Many traditional scientific research fields such as astronomy, fluid dynamics, structural biology, and neurosciences;

- Scientific illustration that can benefit from illustrative visualization techniques for scientific data;

- Personal visualization and visual analytics in which we develop solutions for the general audience.

<div align="center" style="color:red">

COMETE Project-Team

</div>

4. Application Domains

4.1. Security and privacy

Participants: Konstantinos Chatzikokolakis, Catuscia Palamidessi, Ehab Elsalamouny, Marco Stronati, Joris Lamare.

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.

4. Application Domains

4.1. Fuel saving by optimizing airplanes trajectories

We have a collaboration with the startup Safety Line on the optimization of trajectories for civil aircrafts. Key points include the reliable identification of the plane parameters (aerodynamic and thrust models) using data from the flight recorders, and the robust trajectory optimization of the climbing and cruise phases.

4.2. Hybrid vehicles

We have a collaboration with IFPEN on the energy management for hybrid vehicles. A significant direction is the analysis and classification of traffic data.

4.3. Energy production planning

We work with colleagues from U. Chile, in the framework of Inria Chile, on the management of electricity production and storage for a microgrid.

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

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.

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

4. Application Domains

4.1. Creative industries

We work closely with creative professionals in the arts and in design, including music composers, musicians, and sound engineers; painters and illustrators; dancers and choreographers; theater groups; graphic and industrial designers; and architects.

4.2. Scientific research

We work with creative professionals in the sciences and engineering, including neuroscientists and doctors; programmers and statisticians; chemists and astrophysicists; and researchers in fluid mechanics.

GALEN Project-Team

4. Application Domains

4.1. Lung Tumor Detection and Characterization

The use of Diffusion Weighted MR Imaging (DWI) is investigated as an alternative tool to radiologists for tumor detection, tumor characterization, distinguishing tumor tissue from non-tumor tissue, and monitoring and predicting treatment response. In collaboration with Hôpitaux Universitaires Henri-Mondor in Paris, France and Chang Gung Memorial Hospital – Linkou in Taipei, Taiwan we investigate the use of model-based methods of 3D image registration, clustering and segmentation towards the development of a framework for automatic interpretation of images, and in particular extraction of meaningful biomarkers in aggressive lymphomas.

4.2. Co-segmentation and Co-registration of Subcortical Brain Structures

New algorithms to perform co-segmentation and co-registration of subcortical brain structures on MRI images are investigated in collaboration with Ecole Polytechnique de Montreal and the Sainte-Justine Hospital Research Center from Montreal. Brain subcortical structures are involved in different neurodegenerative and neuropsychiatric disorders, including schizophrenia, Alzheimers disease, attention deficit, and subtypes of epilepsy. Segmenting these parts of the brain enables a physician to extract indicators, facilitating their quantitative analysis and characterization. We are investigating how estimated maps of semantic labels (obtained using machine learning techniques) can be used as a surrogate for unlabelled data. We are exploring how to combine them with multi-population deformable registration to improve both alignment and segmentation of these challenging brain structures.

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 [50]). 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 [32],

- bounds on the controllability time [28],

- STIRAP processes [99],

- simultaneous control [72],

- optimal control ([68], [41], [52]),

- 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, [33], [69], [90], [47]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [37], [48], [65], [38]). 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 [39], [40] 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], [44].

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 [36], [49]. 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 [35]), 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 [42], [43], [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], [29].

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 [45].

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]).

<p align="center">**GEOMETRICA Project-Team**</p>

4. Application Domains

4.1. Main Application Domains

Our work is mostly of a fundamental nature but finds applications in a variety of application domains. Transfer is mostly conducted via GeometryFactory, the startup company that commercializes CGAL (see Section 8.1.2).

- Medical Imaging
- Numerical simulation
- Geometric modeling
- Visualization
- Data analysis

4.2. Secondary Application Domains

- Geographic information systems
- Geophysics
- Astrophysics
- Material physics

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

4. Application Domains

4.1. Cryptography and Cryptanalysis

In the twenty-first century, cryptography plays two essential roles: 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;
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.

F. Morain and B. 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, F. Morain' elliptic curve point counting and primality proving algorithms are essential tools in the everyday construction of strong public-key cryptosystems, while B. 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.

D. Augot, F. Levy-dit-Vehel, and A. 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, A. Couvreur's work on filtration attacks on codes has an important impact on the design of code-based systems using wild Goppa codes or algebraic geometry codes, and on the choice of parameter sizes for secure implementations.

Coding theory also has important practical applications in the improvement of conventional symmetric cryptosystems. For example, D. Augot's recent work on MDS matrices via BCH codes gives a more efficient construction of optimal diffusion layers in block ciphers. Here we use combinatorial, non-algorithmic properties of codes, in the internals of designs of block ciphers.

While coding theory brings tools as above for the classical problems of encryption, authentication, and so on, it can also provide solutions to new cryptographic problems. This is classically illustrated by the use of Reed-Solomon codes in secret sharing schemes. Grace is involved in the study, construction and implementation of locally decodable codes, which have applications in quite a few cryptographic protocols : *Private Information Retrieval*, *Proofs of Retrievability*, *Proofs of Ownership*, etc.

ILDA Team

4. Application Domains

4.1. Mission-critical systems

Mission-critical contexts of use include emergency response & management, and critical infrastructure operations, such as public transportation systems, communications and power distribution networks, or the operations of large scientific instruments such as particle accelerators and astronomical observatories. Central to these contexts of work is the notion of situation awareness [21], i.e., how workers perceive and understand elements of the environment with respect to time and space, such as maps and geolocated data feeds from the field, and how they form mental models that help them predict future states of those elements. One of the main challenges is how to best assist subject-matter experts in constructing correct mental models and making informed decisions, often under time pressure. This can be achieved by providing them with, or helping them efficiently identify and correlate, relevant and timely information extracted from large amounts of raw data, taking into account the often cooperative nature of their work and the need for task coordination. With this application area, our goal is to investigate novel ways of interacting with computing systems that improve collaborative data analysis capabilities and decision support assistance in a mission-critical, often time-constrained, work context.

4.2. Exploratory analysis of scientific data

Many scientific disciplines are increasingly data-driven, including astronomy, molecular biology, particle physics, or neuroanatomy. While making the right decision under time pressure is often less of critical issue when analyzing scientific data, at least not on the same temporal scale as truly time-critical systems, scientists are still faced with large-to-huge amounts of data. No matter their origin (experiments, remote observations, large-scale simulations), these data are difficult to understand and analyze in depth because of their sheer size and complexity. Challenges include how to help scientists freely-yet-efficiently explore their data, keep a trace of the multiple data processing paths they considered to verify their hypotheses and make it easy to backtrack, and how to relate observations made on different parts of the data and insights gained at different moments during the exploration process. With this application area, our goal is to investigate how data-centric interactive systems can improve collaborative scientific data exploration, where users' goals are more open-ended, and where roles, collaboration and coordination patterns [40] differ from those observed in mission-critical contexts of work.

INFINE Team (section vide)

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 [10]. 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 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 [125], pour des réseaux ferroviaires [76] ou routiers, et l'évaluation de performance des réseaux de communication [67]. 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" [71], [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 [142], [139], [105].

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 [125] 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 [67]. 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 [71], [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 [142], [139], [105].

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 [168]. Ensuite, le rôle de l'algèbre max-plus dans les problèmes d'ordonnancement est bien connu depuis les années 60, les dates de complétion pouvant souvent être calculées à partir d'équations linéaires max-plus. Plus récemment, des représentations de problèmes d'ordonnancement ont pu être obtenues à partir de semi-groupes de matrices max-plus : une première représentation a été obtenue dans [113] pour le cas du "jobshop", une représentation plus simple a été obtenue dans [136] dans le cas du "flowshop". Ce point de vue algébrique a été très utile dans le cas du "flowshop" : il permet de retrouver des résultats anciens de dominance et d'obtenir ainsi de nouvelles bornes [136]. Finalement, en regardant l'algèbre max-plus comme une limite de l'algèbre classique, on peut utiliser des outils algébriques en optimisation combinatoire [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 [168]. Next, the role of max-plus algebra in scheduling problems has been known since the sixties: completion dates can often be computed by max-plus linear equations. Recently, representations of certain scheduling problems using max-plus matrix semigroups have appeared, a first representation was given in [113] for the jobshop case, a simpler representation was given in [136] in the flowshop case. This algebraic point of view turned out to be particularly fruitful in the flowshop case: it allows one to recover old dominance results and to obtain new bounds [136]. Finally, viewing max-plus algebra as a limit of classical algebra allows to use algebraic tools in combinatorial optimisation [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
    }
```

$$x_1 = [0, 0]$$
$$x_2 =]-\infty, 99] \cap (x_1 \cup x_3)$$
$$x_3 = x_2 + [1, 1]$$
$$x_4 = [100, +\infty[\cap (x_1 \cup x_3)$$

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], [110], entre les problèmes de jeux à somme nulle et les problèmes d'analyse statique, qui peut se résumer par le dictionnaire suivant:

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

Pour que le nombre d'états du jeu soit fini, il est nécessaire de se limiter à des treillis d'ensembles ayant un nombre fini de degrés de liberté, ce qui est le cas de domaines communément utilisés (intervalles, ensembles définis par des contraintes de potentiel de type $x_i - x_j \leq$ cst, mais aussi, les "templates" qui sont des sous-classes de polyèdres introduits récemment par Sankaranarayanan, Sipma et Manna [160]). 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 [111], [83], [84],[7]. Une version assez générale de cet algorithme, adaptée au domaine des templates, est décrite dans [110] et a fait l'objet d'une implémentation prototype. Chaque itération combine de la programmation linéaire et des algorithmes de graphes. Des résultats expérimentaux ont montré le caractère effectif de la méthode, avec souvent un gain en précision par rapport aux approches classiques, par exemple pour des programmes comprenant des boucles imbriquées.

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

English version

The abstract interpretation method introduced by P. and R. Cousot [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], [110], between the zero-sum game problems and the static analysis problems, which can be summarized by the following dictionnary:

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

For the game to have a finite state space, we must restrict our attention to lattices of sets with a finite number of degrees of freedom, which is the case of the domains commonly used in static analysis (intervals, sets defined by potentials constraints of the form $x_i - x_j \leq$ cst, and also the subclasses of polyhedra called "templates", introduced recently by Sankaranarayanan, Sipma and Manna [160]). 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 [111], [83], [84],[7]. A rather general version of this policy iteration algorithm, adapted to the domain of templates, is described in [110], together with a prototype implementation. Every iteration combines linear programming and combinatorial algorithms. Some experimental results indicate that the method often leads to invariants which are more accurate than the ones obtained by alternative methods, in particular for some programs with nested loops.

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

4.5. Autres applications/Other applications

L'algèbre max-plus apparaît de manière naturelle dans le calcul de scores de similitudes dans la comparaison de séquences génétiques. Voir par exemple [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 Projecton self-aware networks, and will be searching new cooperations.

4.2. Transport Systems

Participants: Stefan Haar, Serge Haddad, Yann Duplouy, Simon Theissing.

We participate in the IRT *System X*'s *system of systems* program *TMM*, in two projects:

- project MIC on multi-modal transport systems 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.

- the project SVA (*Simulation pour la Sécurité du Véhicule Autonome*), where the PhD Thesis of Yann Duplouy targets the application of formal methods to the development of embedded systems for autonomous vehicles.

4.3. Biological Systems

Participants: Stefan Haar, Serge Haddad, Stefan Schwoon, Thomas Chatain.

We have begun in 2014 to examine concurrency issues in systems biology, and are currently enlarging the scope of our research's applications in this direction. To see the context, note that in recent years, a considerable shift of biologists' interest can be observed, from the mapping of *static* genotypes to *gene expression*, i.e. the processes in which genetic information is used in producing functional products. These processes are far from being uniquely determined by the gene itself, or even jointly with static properties of the environment; rather, *regulation* occurs throughout the expression processes, with specific mechanisms increasing or decreasing the production of various products, and thus modulating the outcome. These regulations are central in understanding cell fate (how does the cell differenciate ? Do mutations occur ? etc), and progress there hinges on our capacity to analyse, predict, monitor and control complex and variegated processes. Our first step in this domain is related in the conference contribution [57], where we apply Petri net unfolding techniques for the efficient computation of *attractors* in a regulatory network; that is, to identify strongly connected reachability components that correspond to stable evolutions, e.g. of a cell that differentiates into a specific functionality (or mutation). This constitutes the starting point of a broader research with Petri net unfolding techniques in regulation. In fact, ,he use of *ordinary* Petri nets for capturing regulatory network (RN) dynamics overcomes the limitations of traditional RN models : those impose e.g. Monotonicity properties in the influence that one factor had upon another, i.e. always increasing or always decreasing, and were thus unable to cover all actual behaviours (see [75]). Rather, we follow the more refined model of boolean networks of automata, where the local states of the different factors jointly detemine which state transitions are possible. For these connectors, ordinary PNs constitute a first approximation, improving greatly over the literature but leaving room for improvement in terms of introducing more refined logical connectors. Future work thus involves transcending this class of PN models. Via unfoldings, one has access – provided efficient techniques are available – to all behaviours of the model, rather than over-or under-approximations as previously. This opens the way to efficiently searching in particular for determinants of the cell fate : which attractors are reachable from a given stage, and what are the factors that decide in favor of one or the other attractor, etc. The list of potential applications in biology and medicine of such a methodology would be too long to reproduce here.

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

4. Application Domains

4.1. Social Networks

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 8.1.1) as well as in the Structured, Social, and Semantic Search project (Section 8.1.2).

4.2. Computational Journalism

Modern journalism increasingly relies on content management technologies in order to represent, store, and query source data and media objects themselves. Writing news articles increasingly requires consulting several sources, interpreting their findings in context, and crossing links between related sources of information. OAKresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. This work will be funded by the ANR ContentCheck project, and a Google Award on Even Thread Extraction. We work in collaboration with Le Monde's "Les Décodeurs" team to investigate these topics.

4.3. Open Data Intelligence

The Web is a vast source of information, to which more is added every day either in unstructured form (Web pages) or, increasingly, as partially structured sources of information, in particular as Open Data sets, which can be seen as connected graphs of data, most frequently described in the RDF data format recommended by the W3C. Further, RDF data is also the most appropriate format for representing structured information extracted automatically from Web pages, such as the DBPedia database extracted from Wikipedia or Google's InfoBoxes. We work on this topic within the 4-year project ODIN started in 2014.

4.4. Hybrid Data Warehousing

Increasingly many modern applications need to exploit data from a variety of formats, including relations, text, trees, graphs etc. The recent development of data management systems aimed at "Big Data", including NoSQL platforms, large-scale distributed systems etc. provides enteprise architects with many systems to chose from. This makes it hard to decide which part of the application data to handle in which system, especially given that each system is best at handling a specific kind of data and a certain class of operations. OAKinvestigates principled techniques for distributing an application's data sources across a variety of systems and data models, based on materialized views. We test our ideas in this area within the Datalyse project.

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

4. Application Domains

4.1. Human neuroimaging data and their use

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

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

- Diffusion-weighted MRI measures the local diffusion of water molecules in the brain at the resolution of 1 to 2mm, in a set of directions (60 typically). Local anisotropy, observed in white matter, yields a local model of fiber orientation that can be integrated into a geometric model of fiber tracts along which water diffusion occurs, and thus provides information on the connectivity structure of the brain.

- Functional MRI measures the blood-oxygen-level-dependent (BOLD) contrast that reflects neural activity in the brain, at a spatial resolution of 1.5 to 3mm, and a temporal resolution of about 2s. This yields a spatially resolved image of brain functional networks that can be modulated either by specific cognitive tasks or exhibit spontaneous co-activations.

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

4.2. High-field MRI

High field MRI as performed at NeuroSpin (7T on humans, 11.7T in 2017, 17.6T on rats) brings an improvement over traditional MRI acquisitions at 1.5T or 3T, related to to a higher signal-to-noise ratio in the data. Depending on the data and applicative context, this gain in SNR can be traded against spatial resolution improvements, thus helping in getting more detailed views of brain structure and function. This comes at the risk of higher susceptibility distortions of the MRI scans and signal inhomogeneities, that need to be corrected for. Improvements at the acquisition level may come from the use of new coils (such as the 32 channels coil on the 7T at NeuroSpin), as well as the use of multi-band sequences [62].

4.3. Technical challenges for the analysis of neuroimaging data

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

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

PARSIFAL Project-Team

4. Application Domains

4.1. Integrating a model checker and a theorem prover

The goal of combining model checking with inductive and co-inductive theorem in a rather appealing one. The strengths of systems in these two different approaches are strikingly different. A model checker is capable of exploring a finite space automatically: such a tool can repeatedly explores all possible cases of a given a computational space. On the other hand, a theorem prover might be able to prove clever things about a search space. For example, a model checker could attempt to discover whether or not there exists a winning strategy for, say, tic-tac-toe while an inductive theorem prover might be able to prove that if there is a winning strategy for one board then there is a winning strategy for any symmetric version of that board. Of course, the ability to combine proofs from these approaches could drastically reduce the amount of state exploration and verification of proof certificates that are needed to prove the existence of winning strategies.

Our first step to providing an integration of model checking and (inductive) theorem proving was the development of a strong logic, that we call $\mathcal{G}$, which extends intuitionistic logic with notions of least and greatest fixed points. We had developed the proof theory of this logic in earlier papers [4] [63]. We have now recently converted the Bedwyr system so that it formally accepts almost all definitions and theorem statements that are accepted by the inductive theorem prover Abella. Thus, these two systems are proving theorems in the same logic and their results can now be shared.

Bedwyr's tabling mechanism has been extended so that its it can make use of previously proved lemmas. For instance, when trying to prove that some board position has a winning strategy, an available stored lemma can now be used to obtain the result if some symmetric board position is already in the table.

For more about recent progress on providing checkable proof certificates for model checking, see the web site for Bedwyr http://slimmer.gforge.inria.fr/bedwyr/.

4.2. Implementing trusted proof checkers

Traditionally, theorem provers—whether interactive or automatic—are usually monolithic: if any part of a formal development was to be done in a particular theorem prover, then the whole of it would need to be done in that prover. Increasingly, however, formal systems are being developed to integrate the results returned from several, independent and high-performing, specialized provers: see, for example, the integration of Isabelle with an SMT solver [62] as well as the Why3 and ESC/Java systems.

Within the Parsifal team, we have been working on foundational aspects of this multi-prover integration problem. As we have described above, we have been developing a formal framework for defining the semantics of proof evidence. We have also been working on prototype checkers of proof evidence which are capable of executing such formal definitions. The proof definition language described in the papers [59], [58] is currently given an implementation in the λProlog programming language [76]. This initial implementation will be able to serve as a "reference" proof checker: others who are developing proof evidence definitions will be able to use this reference checker to make sure that they are getting their definitions to do what they expect.

Using λProlog as an implementation language has both good and bad points. The good points are that it is rather simple to confirm that the checker is, in fact, sound. The language also supports a rich set of abstracts which make it impossible to interfere with the code of the checker (no injection attacks are possible). On the negative side, however, the performance of our λProlog interpreters is lower than that of specially written checkers and kernels.

4.3. Trustworthy implementations of theorem proving techniques

Instead of integrating different provers by exchanging proof evidence and relying on a backend proof-checker, another approach to integration consists in re-implementing the theorem proving techniques as proof-search strategies, on an architecture that guarantees correctness. Focused systems can serve as the basis of such an architecture, identifying points for choice and backtracking, and providing primitives for the exploration of the search space. These form a trusted *Application Programming Interface* that can be used to program and experiment various proof-search heuristics without worrying about correctness. No proof-checking is needed if one trusts the implementation of the API.

This approach has led to the development of the Psyche engine.

Two major research directions are currently being explored, based on the above:

- The first one is about understanding how to deal with quantifiers in presence of one or more theories: On the one hand, traditional techniques for quantified problems, such as *unification* [47] or *quantifier elimination* are usually designed for either the empty theory or very specific theories. On the other hand, the industrial techniques for combining theories (Nelson-Oppen, Shostak) are designed for quantifier-free problems, and quantifiers are there dealt with incomplete *clause instantiation* methods or *trigger*-based techniques [61]. We are working on making the two approaches compatible.

- The above architecture's modular approach raises the question of how its different modules can safely cooperate (in terms of guaranteed correctness), while some of them are trusted and others are not. The issue is particularly acute if some of the techniques are run concurrently and exchange data at unpredictable times. For this we explore new solutions based on Milner's *LCF* [79]. In [30], we argued that our solutions in particular provide a way to fulfil the "Strategy Challenge for SMT-solving" set by De Moura and Passmore [88].

POEMS Project-Team

4. Application Domains

4.1. Acoustics

Two particular subjects have retained our attention recently.

Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, has been for our team a very challenging topic, which gave rise to a lot of open questions, from the modeling until the numerical approximation of existing models. Our works in this area are partially supported by Airbus Group. The final objective is to reduce the noise radiated by planes.

Musical acoustics constitutes a particularly attractive application. We are concerned by the simulation of musical instruments. The objective is both a better understanding of the behavior of existing instruments and an aid for the manufacturing of new instruments. We have successively considered the timpani, the guitar and the piano. This activity is continuing in the framework of the European Project BATWOMAN.

4.2. Electromagnetism

Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology and electromagnetic compatibility. These areas are still motivating research in computational sciences (large scale computation) and mathematical modeling (derivation of simplified models for multiscale problems). These topics are developed in collaboration with CEA, DGA and ONERA.

Electromagnetic propagation in non classical media opens a wide and unexplored field of research in applied mathematics. This is the case of wave propagation in photonic crystals, metamaterials or magnetized plasmas. Two ANR projects (METAMATH and CHROME) support this research.

Other subjects are motivated by our partnership with CEA-LIST concerning the simulation of non-destructive testing methods: the development of asymptotic models for complex electromagnetic networks and the simulation by integral equations of eddy current phenomena.

4.3. 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. A major application topic has emerged during the past years : the non destructive testing by ultra-sounds which is the main topic of our collaboration with CEA-LIST. On the other hand, we are developing efficient integral equation modelling for geophysical and civil engineering applications (soil-structure interaction for civil engineering, seismology).

POPIX Team

4. Application Domains

4.1. Pharmacometrics

Participants: Marc Lavielle, Raphael Kuate.

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 majorally 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)

POSTALE Team (section vide)

<div align="center">

SELECT Project-Team

</div>

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 methodology to address them. Many of our applications involve contracts with industrial partners, e.g., in reliability, although we also have several academic collaborations, e.g., in genetics and image analysis.

4.2. Curve classification

The field of classification for complex data such as curves, functions, spectra and time series, is an important problem in current research. Standard data analysis questions are being looked into anew, in order to define novel strategies that take the functional nature of such data into account. Functional data analysis addresses a variety of applied problems, including longitudinal studies, analysis of fMRI data, and spectral calibration.

We are focused in particular on unsupervised classification. In addition to standard questions such as the choice of the number of clusters, the norm for measuring the distance between two observations, and vectors for representing clusters, we must also address a major computational problem: the functional nature of the data, which requires new approaches.

4.3. Computer experiments and reliability

For several years now, SELECT has collaborated with the EDF-DER *Maintenance des Risques Industriels* group. One important theme involves 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, a high-technology group (Aerospace propulsion, Aicraft equipment, Defense Security, Communications).

Moreover, a collaboration has begun with Dassault Aviation on the modal analysis of mechanical structures, which aims to identify the vibration behavior of structures under dynamic excitation. From the algorithmic point of view, modal analysis amounts to estimation in parametric models on the basis of measured excitations and structural response data. In literature and existing implementations, the model selection problem associated with this estimation is currently treated by a rather weighty and heuristic procedure. In the context of our own research, model selection via penalization methods are to be tested on this model selection problem.

4.4. Dynamic contrast enhanced imaging

Yves Rozenholc was with SELECT until Septemnber 2015, and introduced research for quantifying tumor microcirculation to monitor treatments in cancer. Dynamic Contrast Enhanced (DCE) imaging provides information on the properties of vascular networks. It enables biostatisticians to design biomarkers that can be used for diagnosis, prognosis and treatment monitoring. To make available robust tumoral microcirculation biomarkers in DCE imaging, Yves Rozenholc has developed several tools for denoising and clustering the dynamics found in DCE imaging sequences, and testing equality of survival functions coming from two DCE imaging sequences.

4.5. Analysis of genomic data

For many years now, 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 has started a thesis co-supervised 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.

SELECT collaborates with Anavaj Sakuntabhai and Benno Schwikowski (Pasteur Institute) on prediction of dengue fever severity from high-dimensional gene expression data. One project involves using/finding new and computationally efficient methods (e.g., 2d isotonic regression, lasso regression) for predicting dengue severity. Due to the high-dimensional nature of the data and low-dimensional nature of the number of individuals, false discovery rate (FDR) methods are used to provide statistical justification of results. A second project involves statistical meta-analysis of newly collected dengue gene expression data along with recently published data sets from other groups.

SELECT is involved in the ANR "jeunes chercheurs" MixStatSeq directed by Cathy Maugis (INSA Toulouse), which is concerned with statistical analysis and clustering of RNASeq genomics data.

4.6. Pharmacovigilance

A collaboration has been started with Pascale Tubert-Bitter, Ismael Ahmed and Mohamed Sedki (Pharmacoepidemiology and Infectious Diseases, PhEMI) for the analysis of pharmacovigilance data. In this framework, the goal is to detect, as soon as possible, potential associations between certain drugs and adverse effects, which appeared after the authorized marketing of these drugs. Instead of working on aggregate data (contingency table) like is usually the case, the approach developed aims to deal with individual's data, which perhaps gives more information. Valerie Robert has begun a thesis co-supervised by Gilles Celeux and Christine Keribin on this topic, which should enable the development of a new model-based clustering method, inspired by latent block models.

4.7. Spectroscopic imaging analysis of ancient materials

Ancient materials, encountered in archaeology and paleontology are often complex, heterogeneous and poorly characterized before physico-chemical analysis. A popular technique to gather as much physico-chemical information as possible, is spectro-microscopy or spectral imaging, where a full spectra, made of more than a 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 requires the combination of an "image" approach with a "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 datasets. This technique enables the simultaneous accounting for spatial and spectral information, while producing statistically sound information on morphological and physico-chemical aspects of the studied samples.

SPECFUN Project-Team (section vide)

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);
  - decades, for investments.

Significant challenges also include:

- spatial distribution of problems; due to capacity limits we can not consider a power grid like Europe + North Africa as a single "production = demand" constraint; with extra connections we can equilibrate excess production by renewables for remote areas, but not in an unlimited manner.
- other uncertainties, which might be modelized by adversarial or stochastic frameworks (e.g. technological breakthroughs, decisions about ecological penalization).

We have had several related projects (Citines, a European (FP7) project; IOMCA, a ANR project), and we now work on the POST project, a ADEME BIA about investments in power systems. Our collaboration with company Artelys (working on optimization in general, and in particular on energy management) is formalized as an Inria ILAB.

Technical challenges: Our work focuses on the combination of reinforcement learning tools, with their anytime behavior and asymptotic guarantees, with existing fast approximate algorithms. 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. We study various modelling errors, such as bias due to finite samples, linearization, and propose corrections.

Related Activities:

- Joint team with Taiwan, namely the Indema associate team (see Section 9.4.1.1).
- Ilab METIS, in progress with Artelys (see Section 6.6) for industrialization of our work. In particular, the Crystal tool is adopted by the European Community (http://www.artelys.com/news/120/90/Energy-The-European-Commission-Chooses-Artelys-Crystal)
- Organization of various forums and meetings around Energy Management
- Visit of Edgar Galvan Lopez also includes applications to energy management, more precisely Demand-Side Management systems. In [40], Differential Evolution is used to generate optimal plans to use the accumulators of electrical vehicles in order to reduce the peak household consumption loads.

4.2. TAO & Humanities

Several projects related to research for Humanities and/or research transfer have started in 2015:

- Personal semantics (Gregory Grefenstette). In the current digital world, individuals generate increasing amount of personal data. Our work involves discovering semantic axes for organizing and exploiting this data for personal use.

- Gregorius (Cécile Germain & Gregory Grefenstette). An application of semantic structuring and automatic enrichment of existing digital humanities archives.

- Cartolabe (Ph. Caillou, Gregory Grefenstette, Jean-Daniel Fekete - AVIZ, Michèle Sebag). The Cartolabe project applies machine learning techniques to determine comprehensible structures in unstructured data. The goal is to use raw textual data, and underspecified ontologies, to provide intuitive access to pertinent research activities in a large research organisation.

- AmiQap (Ph. Caillou, Michèle Sebag). The multivariate analysis of questionaire data relative to the quality of life at work, in relation with the socio-economical indicators of firms, aims at investigating the relationship between quality of life and economic performances (depending on the activity domain).

- Collaborative Hiring (Ph. Caillou, Michèle Sebag). Thomas Schmitt's PhD, started in 2014, aims at handling job offers and demands matching as a collaborative filtering problem.

- IODS (Wikidata for Science).

Significant challenges include some Big Data problems:

- learning interpretable clusters from bottom-up treatment of mixed text and numerical data

- aligning bottom-up clusters with existing manually created top-down structures

- building a unified system integrating the "dire d'experts".

Partners:

- Amiqap is funded by ISN, with Telecom SES, RITM and La Fabrique as partners.

- The collaborative hiring study is conducted in cooperation with J.P. Nadal from EHESS.

<p style="text-align:center"><strong style="color:red">TOCCATA Project-Team</p>

4. Application Domains

4.1. Safety-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. Currently our industrial collaborations mainly belong to the domain of transportation, including aeronautics, railroad, space flight, automotive.

Verification of C programs, Alt-Ergo at Airbus   Transportation is the domain considered in the context of the ANR U3CAT project, led by CEA, in partnership with Airbus France, Dassault Aviation, Sagem Défense et Sécurité. It included proof of C programs via Frama-C/Jessie/Why, proof of floating-point programs [99], the use of the Alt-Ergo prover via CAVEAT tool (CEA) or Frama-C/WP. Within this context, we contributed to a qualification process of Alt-Ergo with Airbus industry: the technical documents (functional specifications and benchmark suite) have been accepted by Airbus, and these documents were submitted by Airbus to the certification authorities (DO-178B standard) in 2012. This action is continued in the new project Soprano.

Certified compilation, certified static analyzers   Aeronautics is the main target of the Verasco project, led by Verimag, on the development of certified static analyzers, in partnership with Airbus. This is a follow-up of the transfer of the CompCert certified compiler (Inria team Gallium) to which we contributed to the support of floating-point computations [13].

Transfer to the community of Ada development   The former FUI project Hi-Lite, led by Adacore company, introduced the use of Why3 and Alt-Ergo as back-end to SPARK2014, an environment for verification of Ada programs. This is applied to the domain of aerospace (Thales, EADS Astrium). At the very beginning of that project, Alt-Ergo was added in the Spark Pro toolset (predecessor of SPARK2014), developed by Altran-Praxis: Alt-Ergo can be used by customers as an alternate prover for automatically proving verification conditions. Its usage is described in the new edition of the Spark book [0] (Chapter "Advanced proof tools"). This action is continued in the new joint laboratory ProofInUse. A recent paper [60] provides an extensive list of applications of SPARK, a major one being the British air control management *iFacts*.

Transfer to the community of Atelier B   In the current ANR project BWare, we investigate the use of Why3 and Alt-Ergo as an alternative back-end for checking proof obligations generated by *Atelier B*, whose main applications are railroad-related software [0], a collaboration with Mitsubishi Electric R&D Centre Europe (Rennes) (joint publication [104]) and ClearSy (Aix-en-Provence).

SMT-based Model-Checking: Cubicle   S. Conchon (with A. Mebsout and F. Zaidi from VALS team at LRI) has a long-term collaboration with S. Krstic and A. Goel (Intel Strategic Cad Labs in Hillsboro, OR, USA) that aims in the development of the SMT-based model checker Cubicle (http://cubicle.lri.fr/) based on Alt-Ergo [101][8]. It is particularly targeted to the verification of concurrent programs and protocols.

Apart from transportation, energy is naturally an application in particular with our long-term partner CEA, in the context of U3CAT and Soprano projects. We also indirectly target communications and data, in particular in contexts with a particular need for security or confidentiality: smart phones, Web applications, health records, electronic voting, etc. These are part of the applications of SPARK [60], including verification of security-related properties, including cryptographic algorithms. Also, our new AJACS project addresses issues related to security and privacy in web applications written in Javascript, also including correctness properties.

---

[0]http://www.altran-praxis.com/book/
[0]http://www.methode-b.com/documentation_b/ClearSy-Industrial_Use_of_B.pdf