



RESEARCH CENTER
Sophia Antipolis - Méditerranée

FIELD

Activity Report 2016

Section Scientific Foundations

Edition: 2017-08-25

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

1. AOSTE Project-Team 4
 2. AROMATH Project-Team 8
 3. DATASHAPE Team 10
 4. MARELLE Project-Team 12

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

5. ACUMES Project-Team 13
 6. APICS Project-Team 18
 7. ECUADOR Project-Team 30
 8. MCTAO Project-Team 33
 9. NACHOS Project-Team 36
 10. TOSCA Project-Team 40

DIGITAL HEALTH, BIOLOGY AND EARTH

11. ABS Project-Team 41
 12. ASCLEPIOS Project-Team 45
 13. ATHENA Project-Team 48
 14. BIOCORE Project-Team 52
 15. BIOVISION Team 54
 16. CAMIN Team 57
 17. CASTOR Project-Team 59
 18. COFFEE Project-Team 60
 19. LEMON Team 62
 20. MATHNEURO Team 67
 21. MORPHEME Project-Team 69
 22. VIRTUAL PLANTS Project-Team 71

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

23. COATI Project-Team 73
 24. DIANA Project-Team 74
 25. FOCUS Project-Team 76
 26. INDES Project-Team 77
 27. MAESTRO Project-Team 78

PERCEPTION, COGNITION AND INTERACTION

28. AYIN Team 80
 29. GRAPHDECO Project-Team 82
 30. GRAPHIK Project-Team 85
 31. HEPHAISTOS Project-Team 87
 32. LAGADIC Project-Team 90
 33. STARS Project-Team 93
 34. TITANE Project-Team 99
 35. WIMMICS Project-Team 102
 36. ZENITH Project-Team 104

AOSTE Project-Team

3. Research Program

3.1. Models of Computation and Communication (MoCCs)

Participants: Julien Deantoni, Robert de Simone, Frédéric Mallet, Dumitru Potop Butucaru.

Esterel, SyncCharts, synchronous formalisms, Process Networks, Marked Graphs, Kahn networks, compilation, synthesis, formal verification, optimization, allocation, refinement, scheduling

Formal Models of Computation form the basis of our approach to Embedded System Design. Because of the growing importance of communication handling, it is now associated with the name, MoCC in short. The appeal of MoCCs comes from the fact that they combine features of mathematical models (formal analysis, transformation, and verification) with these of executable specifications (close to code level, simulation, and implementation). Examples of MoCCs in our case are mainly synchronous reactive formalisms and dataflow process networks. Various extensions or specific restrictions enforce respectively greater expressivity or more focused decidable analysis results.

DataFlow Process Networks and Synchronous Reactive Languages such as ESTEREL/SYNCHARTS and SIGNAL/POLYCHRONY [54], [55], [49], [15], [4], [13] share one main characteristics: they are specified in a self-timed or loosely timed fashion, in the asynchronous data-flow style. But formal criteria in their semantics ensure that, under good correctness conditions, a sound synchronous interpretation can be provided, in which all treatments (computations, signaling communications) are precisely temporally mapped. This is referred to as clock calculus in synchronous reactive systems, and leads to a large body of theoretical studies and deep results in the case of DataFlow Process Networks [50], [48] (consider SDF balance equations for instance [56]).

As a result, explicit schedules become an important ingredient of design, which ultimately can be considered and handled by the designer him/herself. In practice such schedules are sought to optimize other parts of the design, mainly buffering queues: production and consumption of data can be regulated in their relative speeds. This was specially taken into account in the recent theories of Latency-Insensitive Design [51], or N-synchronous processes [52], with some of our contributions [6].

Explicit schedule patterns should be pictured in the framework of low-power distributed mapping of embedded applications onto manycore architectures, where they could play an important role as theoretical formal models on which to compute and optimize allocations and performances. We describe below two lines of research in this direction. Striking in these techniques is the fact that they include time and timing as integral parts of early functional design. But this original time is logical, multiform, and only partially ordering the various functional computations and communications. This approach was radically generalized in our team to a methodology for logical time based design, described next (see 3.2).

3.1.1. K-periodic static scheduling and routing in Process Networks

In the recent years we focused on the algorithm treatments of ultimately k-periodic schedule regimes, which are the class of schedules obtained by many of the theories described above. An important breakthrough occurred when realizing that the type of ultimately periodic binary words that were used for reporting *static scheduling* results could also be employed to record a completely distinct notion of ultimately k-periodic route switching patterns, and furthermore that commonalities of representation could ease combine them together. A new model, by the name of K-periodical Routed marked Graphs (KRG) was introduced, and extensively studied for algebraic and algorithmic properties [5].

The computations of optimized static schedules and other optimal buffering configurations in the context of latency-insensitive design led to the K-Passa software tool development (now terminated)

3.1.2. Endochrony and GALS implementation of conflict-free polychronous programs

The possibility of exploring various schedulings for a given application comes from the fact that some behaviors are truly concurrent, and mutually *conflict-free* (so they can be executed independently, with any choice of ordering). Discovering potential asynchronous inside synchronous reactive specifications then becomes something highly desirable. It can benefit to potential distributed implementation, where signal communications are restricted to a minimum, as they usually incur loss in performance and higher power consumption. This general line of research has come to be known as Endochrony, with some of our contributions [11].

3.2. Logical Time in Model-Driven Embedded System Design

Participants: Julien Deantoni, Frédéric Mallet, Marie Agnes Peraldi Frati, Robert de Simone.

Starting from specific needs and opportunities for formal design of embedded systems as learned from our work on MoCCs (see 3.1), we developed a Logical Time Model as part of the official **OMG UML profile MARTE** for Modeling and Analysis of Real-Time Embedded systems. With this model is associated a Clock Constraint Specification Language (CCSL), which allows to provide loose or strict logical time constraints between design ingredients, be them computations, communications, or any kind of events whose repetitions can be conceived as generating a logical conceptual clock (or activation condition). The definition of CCSL is provided in [1].

Our vision is that many (if not all) of the timing constraints generally expressed as physical prescriptions in real-time embedded design (such as periodicity, sporadicity) could be expressed in a logical setting, while actually many physical timing values are still unknown or unspecified at this stage. On the other hand, our logical view may express much more, such as loosely stated timing relations based on partial orderings or partial constraints.

So far we have used CCSL to express important phenomena as present in several formalisms: **AADL** (used in avionics domain), **EAST-ADL2** (proposed for the **AutoSar** automotive electronic design approach), **IP-Xact** (for System-on-Chip (*SoC*) design). The difference here comes from the fact that these formalisms were formerly describing such issues in informal terms, while CCSL provides a dedicated formal mathematical notation. Close connections with synchronous and polychronous languages, especially Signal, were also established; so was the ability of CCSL to model dataflow process network static scheduling.

In principle the MARTE profile and its Logical Time Model can be used with any UML editor supporting profiles. It has also evolved to become a Domain-Specific Language, independent of UML. It is connected to the **CAPELLA** environment, and the **PAPYRUS** open-source editor. We developed under Eclipse the **TIMESQUARE** solver and emulator for CCSL constraints (see 5.6), with its own graphical interface, as a stand-alone software module, again now coupled with MARTE and Papyrus, but also as part of the **GeMoC studio** environment developed in the GeMoC ANR project.

The MARTE profile and its Logical Time Model can be used with any UML editor supporting profiles but evolved to become a DSL independent of UML. We developed as a set of eclipse plugins the TIMESQUARE tool to edit and simulate CCSL specifications. TimeSquare has been coupled with various tools like Papyrus or Capella and is now part of the concurrent solver integrated in the GEMOC studio.

While CCSL constraints may be introduced as part of the intended functionality, some may also be extracted from requirements imposed either from real-time user demands, or from the resource limitations and features from the intended execution platform. Sophisticated detailed descriptions of platform architectures are allowed using MARTE, as well as formal allocations of application operations (computations and communications) onto platform resources (processors and interconnects). This is of course of great value at a time where embedded architectures are becoming more and more heterogeneous and parallel or distributed, so that application mapping in terms of spatial allocation and temporal scheduling becomes harder and harder. This approach is extensively supported by the MARTE profile and its various models. As such it originates from the Application-Architecture-Adequation (AAA) methodology, first proposed by Yves Sorel, member of Aoste. AAA aims at specific distributed real-time algorithmic methods, described next in 3.3 .

Of course, while logical time in design is promoted here, and our works show how many current notions used in real-time and embedded systems synthesis can naturally be phrased in this model, there will be in the end a phase of validation of the logical time assumptions (as is the case in synchronous circuits and SoC design with timing closure issues). This validation is usually conducted from Worst-Case Execution Time (WCET) analysis on individual components, which are then used in further analysis techniques to establish the validity of logical time assumptions (as partial constraints) asserted during the design.

3.3. The AAA (Algorithm-Architecture Adequation) methodology and Real-Time Scheduling

Participants: Liliana Cucu, Laurent George, Dumitru Potop Butucaru, Yves Sorel.

Note: The AAA methodology and the SynDEX environment are fully described at <http://www.syndex.org/>, together with [relevant publications](#).

3.3.1. Algorithm-Architecture Adequation

The **AAA methodology** relies on distributed real-time scheduling and relevant optimization to connect an Algorithm/Application model to an Architectural one. We now describe its premises and benefits.

The Algorithm model is an extension of the well known data-flow model from Dennis [53]. It is a directed acyclic hyper-graph (DAG) that we call “conditioned factorized data dependence graph”, whose vertices are “operations” and hyper-edges are directed “data or control dependences” between operations. The data dependences define a partial order on the operations execution. The basic data-flow model was extended in three directions: first infinite (resp. finite) repetition of a sub-graph pattern in order to specify the reactive aspect of real-time systems (resp. in order to specify the finite repetition of a sub-graph consuming different data similar to a loop in imperative languages), second “state” when data dependences are necessary between different infinite repetitions of the sub-graph pattern introducing cycles which must be avoided by introducing specific vertices called “delays” (similar to z^{-n} in automatic control), third “conditioning” of an operation by a control dependence similar to conditional control structure in imperative languages, allowing the execution of alternative subgraphs. Delays combined with conditioning allow the programmer to specify automata necessary for describing “mode changes”.

The Architecture model is a directed graph, whose vertices are of two types: “processor” (one sequencer of operations and possibly several sequencers of communications) and “medium” (support of communications), and whose edges are directed connections.

The resulting implementation model [9] is obtained by an external compositional law, for which the architecture graph operates on the algorithm graph. Thus, the result of such compositional law is an algorithm graph, “architecture-aware”, corresponding to refinements of the initial algorithm graph, by computing spatial (distribution) and timing (scheduling) allocations of the operations onto the architecture graph resources. In that context “Adequation” refers to some search amongst the solution space of resulting algorithm graphs, labelled by timing characteristics, for one algorithm graph which verifies timing constraints and optimizes some criteria, usually the total execution time and the number of computing resources (but other criteria may exist). The next section describes distributed real-time schedulability analysis and optimization techniques for that purpose.

3.3.2. Distributed Real-Time Scheduling and Optimization

We address two main issues: uniprocessor and multiprocessor real-time scheduling where constraints must mandatorily be met, otherwise dramatic consequences may occur (hard real-time) and where resources must be minimized because of embedded features.

In the case of uniprocessor real-time scheduling, besides the classical deadline constraint, often equal to a period, we take into consideration dependences between tasks and several, latencies. The latter are complex related “end-to-end” constraints. Dealing with multiple real-time constraints raises the complexity of the scheduling problems. Moreover, because the preemption leads, at least, to a waste of resources due to its approximation in the WCET (Worst Execution Time) of every task, as proposed by Liu and Leyland [57], we first studied non-preemptive real-time scheduling with dependences, periodicities, and latencies constraints. Although a bad approximation of the preemption cost, may have dramatic consequences on real-time scheduling, there are only few researches on this topic. We have been investigating preemptive real-time scheduling since few years, and we focus on the exact cost of the preemption. We have integrated this cost in the schedulability conditions that we propose, and in the corresponding scheduling algorithms. More generally, we are interested in integrating in the schedulability analyses the cost of the RTOS (Real-Time Operating System), for which the cost of preemption is the most difficult part because it varies according to the instance (job) of each task.

In the case of multiprocessor real-time scheduling, we chose at the beginning the partitioned approach, rather than the global approach, since the latter allows task migrations whose cost is prohibitive for current commercial processors. The partitioned approach enables us to reuse the results obtained in the uniprocessor case in order to derive solutions for the multiprocessor case. We consider also the semi-partitioned approach which allows only some migrations in order to minimize the overhead they involve. In addition to satisfy the multiple real-time constraints mentioned in the uniprocessor case, we have to minimize the total execution time (makespan) since we deal with automatic control applications involving feedback loops. Furthermore, the domain of embedded systems leads to solving minimization resources problems. Since these optimization problems are NP-hard we develop exact algorithms (B & B, B & C) which are optimal for simple problems, and heuristics which are sub-optimal for realistic problems corresponding to industrial needs. Long time ago we proposed a very fast “greedy” heuristics [8] whose results were regularly improved, and extended with local neighborhood heuristics, or used as initial solutions for metaheuristics.

In addition to the spatial dimension (distributed) of the real-time scheduling problem, other important dimensions are the type of communication mechanisms (shared memory vs. message passing), or the source of control and synchronization (event-driven vs. time-triggered). We explore real-time scheduling on architectures corresponding to all combinations of the above dimensions. This is of particular impact in application domains such as automotive and avionics (see 4.3).

The arrival of complex hardware responding to the increasing demand for computing power in next generation systems exacerbates the limitations of the current worst-case real-time reasoning. Our solution to overcome these limitations is based on the fact that worst-case situations may have a extremely low probability of appearance within one hour of functioning (10^{-45}), compared to the certification requirements for instance (10^{-9} for the highest level of certification in avionics). Thus we model and analyze the real-time systems using probabilistic models and we propose results that are fundamental for the probabilistic worst-case reasoning over a given time window.

AROMATH Project-Team

3. Research Program

3.1. High order geometric modeling

The accurate description of shapes is a long standing problem in mathematics, with an important impact in many domains, inducing strong interactions between geometry and computation. Developing precise geometric modeling techniques is a critical issue in CAD-CAM. Constructing accurate models, that can be exploited in geometric applications, from digital data produced by cameras, laser scanners, observations or simulations is also a major issue in geometry processing. A main challenge is to construct models that can capture the geometry of complex shapes, using few parameters while being precise.

Our first objective is to develop methods, which are able to describe accurately and in an efficient way, objects or phenomena of geometric nature, using algebraic representations.

The approach followed in CAGD, to describe complex geometry is based on parametric representations called NURBS (Non Uniform Rational B-Spline). The models are constructed by trimming and gluing together high order patches of algebraic surfaces. These models are built from the so-called B-Spline functions that encode a piecewise algebraic function with a prescribed regularity at the seams. Although these models have many advantages and have become the standard for designing nowadays CAD models, they also have important drawbacks. Among them, the difficulty to locally refine a NURBS surface and also the topological rigidity of NURBS patches that imposes to use many such patches with trims for designing complex models, with the consequence of the appearing of cracks at the seams. To overcome these difficulties, an active area of research is to look for new blending functions for the representation of CAD models. Some examples are the so-called T-Splines, LR-Spline blending functions, or hierarchical splines, that have been recently devised in order to perform efficiently local refinement. An important problem is to analyze spline spaces associated to general subdivisions, which is of particular interest in higher order Finite Element Methods. Another challenge in geometric modeling is the efficient representation and/or reconstruction of complex objects, and the description of computational domains in numerical simulation. To construct models that can represent efficiently the geometry of complex shapes, we are interested in developing modeling methods, based on alternative constructions such as skeleton-based representations. The change of representation, in particular between parametric and implicit representations, is of particular interest in geometric computations and in its applications in CAGD.

We also plan to investigate adaptive hierarchical techniques, which can locally improve the approximation of a shape or a function. They shall be exploited to transform digital data produced by cameras, laser scanners, observations or simulations into accurate and structured algebraic models.

The precise and efficient representation of shapes also leads to the problem of extracting and exploiting characteristic properties of shapes such as symmetry, which is very frequent in geometry. Reflecting the symmetry of the intended shape in the representation appears as a natural requirement for visual quality, but also as a possible source of sparsity of the representation. Recognizing, encoding and exploiting symmetry requires new paradigms of representation and further algebraic developments. Algebraic foundations for the exploitation of symmetry in the context of non linear differential and polynomial equations are addressed. The intent is to bring this expertise with symmetry to the geometric models and computations developed by AROMATH.

3.2. Robust algebraic-geometric computation

In many problems, digital data are approximated and cannot just be used as if they were exact. In the context of geometric modeling, polynomial equations appear naturally, as a way to describe constraints between the unknown variables of a problem. *An important challenge is to take into account the input error in order to*

develop robust methods for solving these algebraic constraints. Robustness means that a small perturbation of the input should produce a controlled variation of the output, that is forward stability, when the input-output map is regular. In non-regular cases, robustness also means that the output is an exact solution, or the most coherent solution, of a problem with input data in a given neighborhood, that is backward stability.

Our second long term objective is to develop methods to robustly and efficiently solve algebraic problems that occur in geometric modeling.

Robustness is a major issue in geometric modeling and algebraic computation. Classical methods in computer algebra, based on the paradigm of exact computation, cannot be applied directly in this context. They are not designed for stability against input perturbations. New investigations are needed to develop methods, which integrate this additional dimension of the problem. Several approaches are investigated to tackle these difficulties.

One is based on linearization of algebraic problems based on “elimination of variables” or projection into a space of smaller dimension. Resultant theory provides strong foundation for these methods, connecting the geometric properties of the solutions with explicit linear algebra on polynomial vector spaces, for families of polynomial systems (e.g., homogeneous, multi-homogeneous, sparse). Important progresses have been made in the last two decades to extend this theory to new families of problems with specific geometric properties. Additional advances have been achieved more recently to exploit the syzygies between the input equations. This approach provides matrix based representations, which are particularly powerful for approximate geometric computation on parametrized curves and surfaces. They are tuned to certain classes of problems and an important issue is to detect and analyze degeneracies and to adapt them to these cases.

A more adaptive approach involves linear algebra computation in a hierarchy of polynomial vector spaces. It produces a description of quotient algebra structures, from which the solutions of polynomial systems can be recovered. This family of methods includes Gröbner Basis, which provides general tools for solving polynomial equations. Border Basis is an alternative approach, offering numerically stable methods for solving polynomial equations with approximate coefficients. An important issue is to understand and control the numerical behavior of these methods as well as their complexity and to exploit the structure of the input system.

In order to compute “only” the (real) solutions of a polynomial system in a given domain, duality techniques can also be employed. They consist in analyzing and adding constraints on the space of linear forms which vanish on the polynomial equations. Combined with semi-definite programming techniques, they provide efficient methods to compute the real solutions of algebraic equations or to solve polynomial optimization problems. The main issues are the completeness of the approach, their scalability with the degree and dimension and the certification of bounds.

Singular solutions of polynomial systems can be analyzed by computing differentials, which vanish at these points. This leads to efficient deflation techniques, which transform a singular solution of a given problem into a regular solution of the transformed problem. These local methods need to be combined with more global root localisation methods.

Subdivision methods are another type of methods which are interesting for robust geometric computation. They are based on exclusion tests which certify that no solution exists in a domain and inclusion tests, which certify the uniqueness of a solution in a domain. They have shown their strength in addressing many algebraic problems, such as isolating real roots of polynomial equations or computing the topology of algebraic curves and surfaces. The main issues in these approaches is to deal with singularities and degenerate solutions.

DATASHAPE Team

3. Research Program

3.1. Algorithmic aspects of topological and geometric data analysis

TDA requires to construct and manipulate appropriate representations of complex and high dimensional shapes. A major difficulty comes from the fact that the complexity of data structures and algorithms used to approximate shapes rapidly grows as the dimensionality increases, which makes them intractable in high dimensions. We focus our research on simplicial complexes which offer a convenient representation of general shapes and generalize graphs and triangulations. Our work includes the study of simplicial complexes with good approximation properties and the design of compact data structures to represent them.

In low dimensions, effective shape reconstruction techniques exist that can provide precise geometric approximations very efficiently and under reasonable sampling conditions. Extending those techniques to higher dimensions as is required in the context of TDA is problematic since almost all methods in low dimensions rely on the computation of a subdivision of the ambient space. A direct extension of those methods would immediately lead to algorithms whose complexities depend exponentially on the ambient dimension, which is prohibitive in most applications. A first direction to by-pass the curse of dimensionality is to develop algorithms whose complexities depend on the intrinsic dimension of the data (which most of the time is small although unknown) rather than on the dimension of the ambient space. Another direction is to resort to cruder approximations that only captures the homotopy type or the homology of the sampled shape. The recent theory of persistent homology provides a powerful and robust tool to study the homology of sampled spaces in a stable way.

3.2. Statistical aspects of topological and geometric data analysis

The wide variety of larger and larger available data - often corrupted by noise and outliers - requires to consider the statistical properties of their topological and geometric features and to propose new relevant statistical models for their study.

There exist various statistical and machine learning methods intending to uncover the geometric structure of data. Beyond manifold learning and dimensionality reduction approaches that generally do not allow to assert the relevance of the inferred topological and geometric features and are not well-suited for the analysis of complex topological structures, set estimation methods intend to estimate, from random samples, a set around which the data is concentrated. In these methods, that include support and manifold estimation, principal curves/manifolds and their various generalizations to name a few, the estimation problems are usually considered under losses, such as Hausdorff distance or symmetric difference, that are not sensitive to the topology of the estimated sets, preventing these tools to directly infer topological or geometric information.

Regarding purely topological features, the statistical estimation of homology or homotopy type of compact subsets of Euclidean spaces, has only been considered recently, most of the time under the quite restrictive assumption that the data are randomly sampled from smooth manifolds.

In a more general setting, with the emergence of new geometric inference tools based on the study of distance functions and algebraic topology tools such as persistent homology, computational topology has recently seen an important development offering a new set of methods to infer relevant topological and geometric features of data sampled in general metric spaces. The use of these tools remains widely heuristic and until recently there were only a few preliminary results establishing connections between geometric inference, persistent homology and statistics. However, this direction has attracted a lot of attention over the last three years. In particular, stability properties and new representations of persistent homology information have led to very promising results to which the DATASHAPE members have significantly contributed. These preliminary results open many perspectives and research directions that need to be explored.

Our goal is to build on our first statistical results in TDA to develop the mathematical foundations of Statistical Topological and Geometric Data Analysis. Combined with the other objectives, our ultimate goal is to provide a well-founded and effective statistical toolbox for the understanding of topology and geometry of data.

3.3. Topological approach for multimodal data processing

Due to their geometric nature, multimodal data (images, video, 3D shapes, etc.) are of particular interest for the techniques we develop. Our goal is to establish a rigorous framework in which data having different representations can all be processed, mapped and exploited jointly. This requires adapting our tools and sometimes developing entirely new or specialized approaches.

The choice of multimedia data is motivated primarily by the fact that the amount of such data is steadily growing (with e.g. video streaming accounting for nearly two thirds of peak North-American Internet traffic, and almost half a billion images being posted on social networks each day), while at the same time it poses significant challenges in designing informative notions of (dis)-similarity as standard metrics (e.g. Euclidean distances between points) are not relevant.

3.4. Experimental research and software development

We develop a high quality open source software platform called GUDHI which is becoming a reference in geometric and topological data analysis in high dimensions. The goal is not to provide code tailored to the numerous potential applications but rather to provide the central data structures and algorithms that underly applications in geometric and topological data analysis.

The development of the GUDHI platform also serves to benchmark and optimize new algorithmic solutions resulting from our theoretical work. Such development necessitates a whole line of research on software architecture and interface design, heuristics and fine-tuning optimization, robustness and arithmetic issues, and visualization. We aim at providing a full programming environment following the same recipes that made up the success story of the CGAL library, the reference library in computational geometry.

Some of the algorithms implemented on the platform will also be interfaced to other software platform, such as the R software⁰ for statistical computing, and languages such as Python in order to make them usable in combination with other data analysis and machine learning tools. A first attempt in this direction has been done with the creation of an R package called TDA in collaboration with the group of Larry Wasserman at Carnegie Mellon University (Inria Associated team CATS) that already includes some functionalities of the GUDHI library and implements some joint results between our team and the CMU team. A similar interface with the Python language is also considered a priority. To go even further towards helping users, we will provide utilities that perform the most common tasks without requiring any programming at all.

⁰<https://www.r-project.org/>

MARELLE Project-Team

3. Research Program

3.1. Type theory and formalization of mathematics

The calculus of inductive constructions is a branch of type theory that serves as a foundation for theorem proving tools, especially the Coq proof assistant. It is powerful enough to formalize complex mathematics, based on algebraic structures and operations. This is especially important as we want to produce proofs of logical properties for these algebraic structures, a goal that is only marginally addressed in most scientific computation systems.

The calculus of inductive constructions also makes it possible to write algorithms as recursive functional programs which manipulate tree-like data structures. A third important characteristic of this calculus is that it is also a language for manipulating proofs. All this makes this calculus a tool of choice for our investigations. However, this language still is the object of improvements and part of our work focusses on these improvements.

3.2. Verification of scientific algorithms

To produce certified algorithms, we use the following approach: instead of attempting to prove properties of an existing program written in a conventional programming language such as C or Java, we produce new programs in the calculus of constructions whose correctness is an immediate consequence of their construction. This has several advantages. First, we work at a high level of abstraction, independently of the target implementation language. Secondly, we concentrate on specific characteristics of the algorithm, and abstract away from the rest (for instance, we abstract away from memory management or data implementation strategies). Therefore, we are able to address more high-level mathematics and to express more general properties without being overwhelmed by implementation details.

However, this approach also presents a few drawbacks. For instance, the calculus of constructions usually imposes that recursive programs should explicitly terminate for all inputs. For some algorithms, we need to use advanced concepts (for instance, well-founded relations) to make the property of termination explicit, and proofs of correctness become especially difficult in this setting.

3.3. Programming language semantics

To bridge the gap between our high-level descriptions of algorithms and conventional programming languages, we investigate the algorithms that are present in programming language implementations, for instance algorithms that are used in a compiler or a static analysis tool. When working on these algorithms, we usually base our work on the semantic description of the programming language. The properties that we attempt to prove for an algorithm are, for example, that an optimization respects the meaning of programs or that the programs produced are free of some unwanted behavior. In practice, we rely on this study of programming language semantics to propose extensions to theorem proving tools or to verify that compilers for conventional programming languages are exempt from bugs.

ACUMES Project-Team

3. Research Program

3.1. Research directions

The project develops along the following two axes:

- modeling complex systems through novel (unconventional) PDE systems, accounting for multi-scale phenomena and uncertainty;
- optimization and optimal control algorithms for systems governed by the above PDE systems.

These themes are motivated by the specific problems treated in the applications, and represent important and up-to-date issues in engineering sciences. For example, improving the design of transportation means and civil buildings, and the control of traffic flows, would result not only in better performances of the object of the optimization strategy (vehicles, buildings or road networks level of service), but also in enhanced safety and lower energy consumption, contributing to reduce costs and pollutant emissions.

3.1.1. PDE models accounting for multi-scale phenomena and uncertainties

Dynamical models consisting of evolutionary PDEs, mainly of hyperbolic type, appear classically in the applications studied by the previous Project-Team Opale (compressible flows, traffic, cell-dynamics, medicine, etc). Yet, the classical purely macroscopic approach is not able to account for some particular phenomena related to specific interactions occurring at smaller scales. These phenomena can be of greater importance when dealing with particular applications, where the "first order" approximation given by the purely macroscopic approach reveals to be inadequate. We refer for example to self-organizing phenomena observed in pedestrian flows [95], or to the dynamics of turbulent flows for which large scale / small scale vortical structures interfere [123].

Nevertheless, macroscopic models offer well known advantages, namely a sound analytical framework, fast numerical schemes, the presence of a low number of parameters to be calibrated, and efficient optimization procedures. Therefore, we are convinced of the interest of keeping this point of view as dominant, while completing the models with information on the dynamics at the small scale / microscopic level. This can be achieved through several techniques, like hybrid models, homogenization, mean field games. In this project, we will focus on the aspects detailed below.

The development of adapted and efficient numerical schemes is a mandatory completion, and sometimes ingredient, of all the approaches listed below. The numerical schemes developed by the team are based on finite volumes or finite elements techniques, and constitute an important tool in the study of the considered models, providing a necessary step towards the design and implementation of the corresponding optimization algorithms, see Section 3.1.2 .

3.1.1.1. Micro-macro couplings

Modeling of complex problems with a dominant macroscopic point of view often requires couplings with small scale descriptions. Accounting for systems heterogeneity or different degrees of accuracy usually leads to coupled PDE-ODE systems.

In the case of heterogeneous problems the coupling is "intrinsic", i.e. the two models evolve together and mutually affect each-other. For example, accounting for the impact of a large and slow vehicle (like a bus or a truck) on traffic flow leads to a strongly coupled system consisting of a (system of) conservation law(s) coupled with an ODE describing the bus trajectory, which acts as a moving bottleneck. The coupling is realized through a local unilateral moving constraint on the flow at the bus location, see [64] for an existence result and [49], [63] for numerical schemes.

If the coupling is intended to offer higher degree of accuracy at some locations, a macroscopic and a microscopic model are connected through an artificial boundary, and exchange information across it through suitable boundary conditions. See [55], [84] for some applications in traffic flow modelling, and [74], [79], [81] for applications to cell dynamics.

The corresponding numerical schemes are usually based on classical finite volume or finite element methods for the PDE, and Euler or Runge-Kutta schemes for the ODE, coupled in order to take into account the interaction fronts. In particular, the dynamics of the coupling boundaries require an accurate handling capturing the possible presence of non-classical shocks and preventing diffusion, which could produce wrong solutions, see for example [49], [63].

We plan to pursue our activity in this framework, also extending the above mentioned approaches to problems in two or higher space dimensions, to cover applications to crowd dynamics or fluid-structure interaction.

3.1.1.2. Micro-macro limits

Rigorous derivation of macroscopic models from microscopic ones offers a sound basis for the proposed modeling approach, and can provide alternative numerical schemes, see for example [56], [66] for the derivation of Lighthill-Whitham-Richards [107], [122] traffic flow model from Follow-the-Leader and [75] for results on crowd motion models (see also [97]). To tackle this aspect, we will rely mainly on two (interconnected) concepts: measure-valued solutions and mean-field limits.

The notion of **measure-valued solutions** for conservation laws was first introduced by DiPerna [67], and extensively used since then to prove convergence of approximate solutions and deduce existence results, see for example [76] and references therein. Measure-valued functions have been recently advocated as the appropriate notion of solution to tackle problems for which analytical results (such as existence and uniqueness of weak solutions in distributional sense) and numerical convergence are missing [38], [78]. We refer, for example, to the notion of solution for non-hyperbolic systems [86], for which no general theoretical result is available at present, and to the convergence of finite volume schemes for systems of hyperbolic conservation laws in several space dimensions, see [78].

In this framework, we plan to investigate and make use of measure-based PDE models for vehicular and pedestrian traffic flows. Indeed, a modeling approach based on (multi-scale) time-evolving measures (expressing the agents probability distribution in space) has been recently introduced (see the monograph [60]), and proved to be successful for studying emerging self-organised flow patterns [59]. The theoretical measure framework proves to be also relevant in addressing micro-macro limiting procedures of mean field type [87], where one lets the number of agents going to infinity, while keeping the total mass constant. In this case, one must prove that the *empirical measure*, corresponding to the sum of Dirac measures concentrated at the agents positions, converges to a measure-valued solution of the corresponding macroscopic evolution equation. We recall that a key ingredient in this approach is the use of the *Wasserstein distances* [130], [131]. Indeed, as observed in [114], the usual L^1 spaces are not natural in this context, since they don't guarantee uniqueness of solutions.

This procedure can potentially be extended to more complex configurations, like for example road networks or different classes of interacting agents, or to other application domains, like cell-dynamics.

Another powerful tool we shall consider to deal with micro-macro limits is the so-called **Mean Field Games (MFG)** technique (see the seminal paper [106]). This approach has been recently applied to some of the systems studied by the team, such as traffic flow and cell dynamics. In the context of crowd dynamics, including the case of several populations with different targets, the mean field game approach has been adopted in [45], [46], [68], [105], under the assumption that the individual behavior evolves according to a stochastic process, which gives rise to parabolic equations greatly simplifying the analysis of the system. Besides, a deterministic context is studied in [118], which considers a non-local velocity field. For cell dynamics, in order to take into account the fast processes that occur in the migration-related machinery, a framework such the one developed in [62] to handle games "where agents evolve their strategies according to the best-reply scheme on a much faster time scale than their social configuration variables" may turn out to be suitable. An alternative framework to MFG is also considered. This framework is based on the formulation of -Nash- games

constrained by the **Fokker-Planck** (FP, [36]) partial differential equations that govern the time evolution of the probability density functions -PDF- of stochastic systems and on objectives that may require to follow a given PDF trajectory or to minimize an expectation functional.

3.1.1.3. Non-local flows

Non-local interactions can be described through macroscopic models based on integro-differential equations. Systems of the type

$$\partial_t u + \operatorname{div}_{\mathbf{x}} F(t, \mathbf{x}, u, W) = 0, \quad t > 0, \mathbf{x} \in \mathbb{R}^d, d \geq 1, \quad (1)$$

where $u = u(t, \mathbf{x}) \in \mathbb{R}^N$, $N \geq 1$ is the vector of conserved quantities and the variable $W = W(t, x, u)$ depends on an integral evaluation of u , arise in a variety of physical applications. Space-integral terms are considered for example in models for granular flows [33], sedimentation [40], supply chains [89], conveyor belts [90], biological applications like structured populations dynamics [113], or more general problems like gradient constrained equations [34]. Also, non-local in time terms arise in conservation laws with memory, starting from [61]. In particular, equations with non-local flux have been recently introduced in traffic flow modeling to account for the reaction of drivers or pedestrians to the surrounding density of other individuals, see [3], [6] [48], [52], [126]. While pedestrians are likely to react to the presence of people all around them, drivers will mainly adapt their velocity to the downstream traffic, assigning a greater importance to closer vehicles. In particular, and in contrast to classical (without integral terms) macroscopic equations, these models are able to display finite acceleration of vehicles through Lipschitz bounds on the mean velocity [3], [6] and lane formation in crossing pedestrian flows.

General analytical results on non-local conservation laws, proving existence and eventually uniqueness of solutions of the Cauchy problem for (1), can be found in [35] for scalar equations in one space dimension ($N = d = 1$), in [53] for scalar equations in several space dimensions ($N = 1, d \geq 1$) and in [29], [54], [58] for multi-dimensional systems of conservation laws. Besides, specific finite volume numerical methods have been developed recently in [29], [6] and [104].

Relying on these encouraging results, we aim to push a step further the analytical and numerical study of non-local models of type (1), in particular concerning well-posedness of initial - regularity of solutions, boundary value problems and high-order numerical schemes.

3.1.1.4. Uncertainty in parameters and initial-boundary data

Different sources of uncertainty can be identified in PDE models, related to the fact that the problem of interest is not perfectly known. At first, initial and boundary condition values can be uncertain. For instance, in traffic flows, the time-dependent value of inlet and outlet fluxes, as well as the initial distribution of vehicles density, are not perfectly determined [47]. In aerodynamics, inflow conditions like velocity modulus and direction, are subject to fluctuations [93], [112]. For some engineering problems, the geometry of the boundary can also be uncertain, due to structural deformation, mechanical wear or disregard of some details [70]. Another source of uncertainty is related to the value of some parameters in the PDE models. This is typically the case of parameters in turbulence models in fluid mechanics, which have been calibrated according to some reference flows but are not universal [124], [129], or in traffic flow models, which may depend on the type of road, weather conditions, or even the country of interest (due to differences in driving rules and conductors behaviour). This leads to equations with flux functions depending on random parameters [125], [128], for which the mean and the variance of the solutions can be computed using different techniques. Indeed, uncertainty quantification for systems governed by PDEs has become a very active research topic in the last years. Most approaches are embedded in a probabilistic framework and aim at quantifying statistical moments of the PDE solutions, under the assumption that the characteristics of uncertain parameters are known. Note that classical Monte-Carlo approaches exhibit low convergence rate and consequently accurate simulations require huge computational times. In this respect, some enhanced algorithms have been proposed, for example in the balance law framework [111]. Different approaches propose to modify the PDE solvers to account for this probabilistic context, for instance by defining the non-deterministic part of the solution on an orthogonal

basis (Polynomial Chaos decomposition) and using a Galerkin projection [93], [102], [108], [133] or an entropy closure method [65], or by discretizing the probability space and extending the numerical schemes to the stochastic components [28]. Alternatively, some other approaches maintain a fully deterministic PDE resolution, but approximate the solution in the vicinity of the reference parameter values by Taylor series expansions based on first- or second-order sensitivities [119], [129], [132].

Our objective regarding this topic is twofold. In a pure modeling perspective, we aim at including uncertainty quantification in models calibration and validation for predictive use. In this case, the choice of the techniques will depend on the specific problem considered [39]. Besides, we plan to extend previous works on sensitivity analysis [70], [109] to more complex and more demanding problems. In particular, high-order Taylor expansions of the solution (greater than two) will be considered in the framework of the Sensitivity Equation Method [41] (SEM) for unsteady aerodynamic applications, to improve the accuracy of mean and variance estimations. A second targeted topic in this context is the study of the uncertainty related to turbulence closure parameters, in the sequel of [129]. We aim at exploring the capability of the SEM approach to detect a change of flow topology, in case of detached flows. Our ambition is to contribute to the emergence of a new generation of simulation tools, which will provide solution densities rather than values, to tackle real-life uncertain problems. This task will also include a reflection about numerical schemes used to solve PDE systems, in the perspective of constructing a unified numerical framework able to account for exact geometries (isogeometric methods), uncertainty propagation and sensitivity analysis w.r.t. control parameters.

3.1.2. Optimization and control algorithms for systems governed by PDEs

The non-classical models described above are developed in the perspective of design improvement for real-life applications. Therefore, control and optimization algorithms are also developed in conjunction with these models. The focus here is on the methodological development and analysis of optimization algorithms for PDE systems in general, keeping in mind the application domains in the way the problems are mathematically formulated.

3.1.2.1. Sensitivity VS adjoint equation

Adjoint methods (achieved at continuous or discrete level) are now commonly used in industry for steady PDE problems. Our recent developments [121] have shown that the (discrete) adjoint method can be efficiently applied to cost gradient computations for time-evolving traffic flow on networks, thanks to the special structure of the associated linear systems and the underlying one dimensionality of the problem. However, this strategy is questionable for more complex (e.g. 2D/3D) unsteady problems, because it requires sophisticated and time-consuming check-pointing and/or re-computing strategies [37], [88] for the backward time integration of the adjoint variables. The sensitivity equation method (SEM) offers a promising alternative [69], [98], if the number of design parameters is moderate. Moreover, this approach can be employed for other goals, like fast evaluation of neighboring solutions or uncertainty propagation [70].

Regarding this topic, we intend to apply the continuous sensitivity equation method to challenging problems. In particular, in aerodynamics, multi-scale turbulence models like Large-Eddy Simulation (LES) [123], Detached-Eddy Simulation (DES) [127] or Organized-Eddy Simulation (OES) [43], are more and more employed to analyse the unsteady dynamics of the flows around bluff-bodies, because they have the ability to compute the interactions of vortices at different scales, contrary to classical Reynolds-Averaged Navier-Stokes models. However, their use in design optimization is tedious, due to the long time integration required. In collaboration with turbulence specialists (M. Braza, CNRS - IMFT), we aim at developing numerical methods for effective sensitivity analysis in this context, and apply them to realistic problems, like the optimization of active flow control devices. Note that the use of SEM allows computing cost functional gradients at any time, which permits to construct new gradient-based optimization strategies like instantaneous-feedback method [100] or multiobjective optimization algorithm (see section below).

3.1.2.2. Multi-objective descent algorithms for multi-disciplinary, multi-point, unsteady optimization or robust-design

n differentiable optimization, multi-disciplinary, multi-point, unsteady optimization or robust-design can all be formulated as multi-objective optimization problems. In this area, we have proposed the *Multiple-Gradient*

Descent Algorithm (MGDA) to handle all criteria concurrently [71] [72]. Originally, we have stated a principle according which, given a family of local gradients, a descent direction common to all considered objective-functions simultaneously is identified, assuming the Pareto-stationarity condition is not satisfied. When the family is linearly-independent, we dispose of a direct algorithm. Inversely, when the family is linearly-dependent, a quadratic-programming problem should be solved. Hence, the technical difficulty is mostly conditioned by the number m of objective functions relative to the search space dimension n . In this respect, the basic algorithm has recently been revised [73] to handle the case where $m > n$, and even $m \gg n$, and is currently being tested on a test-case of robust design subject to a periodic time-dependent Navier-Stokes flow. The multi-point situation is very similar and, being of great importance for engineering applications, will be treated at large.

Moreover, we intend to develop and test a new methodology for robust design that will include uncertainty effects. More precisely, we propose to employ MGDA to achieve an effective improvement of all criteria simultaneously, which can be of statistical nature or discrete functional values evaluated in confidence intervals of parameters. Some recent results obtained at ONERA [116] by a stochastic variant of our methodology confirm the viability of the approach. A PhD thesis has also been launched at ONERA/DADS.

Lastly, we note that in situations where gradients are difficult to evaluate, the method can be assisted by a meta-model [135].

3.1.2.3. Bayesian Optimization algorithms for efficient computation of general equilibria

Bayesian Optimization -BO- relies on Gaussian processes, which are used as emulators (or surrogates) of the black-box model outputs based on a small set of model evaluations. Posterior distributions provided by the Gaussian process are used to design acquisition functions that guide sequential search strategies that balance between exploration and exploitation. Such approaches have been transposed to frameworks other than optimization, such as uncertainty quantification. Our aim is to investigate how the BO apparatus can be applied to the search of general game equilibria, and in particular the classical Nash equilibrium (NE). To this end, we propose two complementary acquisition functions, one based on a greedy search approach and one based on the Stepwise Uncertainty Reduction paradigm [80]. Our proposal is designed to tackle derivative-free, expensive models, hence requiring very few model evaluations to converge to the solution.

3.1.2.4. Decentralized strategies for inverse problems

Most if not all the mathematical formulations of inverse problems (a.k.a. reconstruction, identification, data recovery, non destructive engineering,...) are known to be ill posed in the Hadamard sense. Indeed, in general, inverse problems try to fulfill (minimize) two or more very antagonistic criteria. One classical example is the Tikhonov regularization, trying to find artificially smoothed solutions close to naturally non-smooth data.

We consider here the theoretical general framework of parameter identification coupled to (missing) data recovery. Our aim is to design, study and implement algorithms derived within a game theoretic framework, which are able to find, with computational efficiency, equilibria between the "identification related players" and the "data recovery players". These two parts are known to pose many challenges, from a theoretical point of view, like the identifiability issue, and from a numerical one, like convergence, stability and robustness problems. These questions are tricky [30] and still completely open for systems like e.g. coupled heat and thermoelastic joint data and material detection.

APICS Project-Team

3. Research Program

3.1. Introduction

Within the extensive field of inverse problems, much of the research by Apics deals with reconstructing solutions of classical elliptic PDEs from their boundary behavior. Perhaps the simplest example lies with harmonic identification of a stable linear dynamical system: the transfer-function f can be evaluated at a point $i\omega$ of the imaginary axis from the response to a periodic input at frequency ω . Since f is holomorphic in the right half-plane, it satisfies there the Cauchy-Riemann equation $\bar{\partial}f = 0$, and recovering f amounts to solve a Dirichlet problem which can be done in principle using, *e.g.* the Cauchy formula.

Practice is not nearly as simple, for f is only measured pointwise in the pass-band of the system which makes the problem ill-posed [70]. Moreover, the transfer function is usually sought in specific form, displaying the necessary physical parameters for control and design. For instance if f is rational of degree n , then $\bar{\partial}f = \sum_1^n a_j \delta_{z_j}$ where the z_j are its poles and δ_{z_j} is a Dirac unit mass at z_j . Thus, to find the domain of holomorphy (*i.e.* to locate the z_j) amounts to solve a (degenerate) free-boundary inverse problem, this time on the left half-plane. To address such questions, the team has developed a two-step approach as follows.

Step 1: To determine a complete model, that is, one which is defined at every frequency, in a sufficiently versatile function class (*e.g.* Hardy spaces). This ill-posed issue requires regularization, for instance constraints on the behavior at non-measured frequencies.

Step 2: To compute a reduced order model. This typically consists of rational approximation of the complete model obtained in step 1, or phase-shift thereof to account for delays. We emphasize that deriving a complete model in step 1 is crucial to achieve stability of the reduced model in step 2.

Step 1 relates to extremal problems and analytic operator theory, see Section 3.3.1 . Step 2 involves optimization, and some Schur analysis to parametrize transfer matrices of given Mc-Millan degree when dealing with systems having several inputs and outputs, see Section 3.3.2.2 . It also makes contact with the topology of rational functions, in particular to count critical points and to derive bounds, see Section 3.3.2 . Step 2 raises further issues in approximation theory regarding the rate of convergence and the extent to which singularities of the approximant (*i.e.* its poles) tend to singularities of the approximated function; this is where logarithmic potential theory becomes instrumental, see Section 3.3.3 .

Applying a realization procedure to the result of step 2 yields an identification procedure from incomplete frequency data which was first demonstrated in [76] to tune resonant microwave filters. Harmonic identification of nonlinear systems around a stable equilibrium can also be envisaged by combining the previous steps with exact linearization techniques from [33].

A similar path can be taken to approach design problems in the frequency domain, replacing the measured behavior by some desired behavior. However, describing achievable responses in terms of the design parameters is often cumbersome, and most constructive techniques rely on specific criteria adapted to the physics of the problem. This is especially true of filters, the design of which traditionally appeals to polynomial extremal problems [72], [56]. Apics contributed to this area the use of Zolotarev-like problems for multi-band synthesis, although we presently favor interpolation techniques in which parameters arise in a more transparent manner, see Section 3.2.2 .

The previous example of harmonic identification quickly suggests a generalization of itself. Indeed, on identifying \mathbb{C} with \mathbb{R}^2 , holomorphic functions become conjugate-gradients of harmonic functions, so that harmonic identification is, after all, a special case of a classical issue: to recover a harmonic function on a domain from partial knowledge of the Dirichlet-Neumann data; when the portion of boundary where data are not available is itself unknown, we meet a free boundary problem. This framework for 2-D non-destructive control was first advocated in [61] and subsequently received considerable attention. It makes clear how to

state similar problems in higher dimensions and for more general operators than the Laplacian, provided solutions are essentially determined by the trace of their gradient on part of the boundary which is the case for elliptic equations⁰ [13], [79]. Such questions are particular instances of the so-called inverse potential problem, where a measure μ has to be recovered from the knowledge of the gradient of its potential (*i.e.*, the field) on part of a hypersurface (a curve in 2-D) encompassing the support of μ . For Laplace's operator, potentials are logarithmic in 2-D and Newtonian in higher dimensions. For elliptic operators with non constant coefficients, the potential depends on the form of fundamental solutions and is less manageable because it is no longer of convolution type. Nevertheless it is a useful concept bringing perspective on how problems could be raised and solved, using tools from harmonic analysis.

Inverse potential problems are severely indeterminate because infinitely many measures within an open set produce the same field outside this set; this phenomenon is called *balayage* [69]. In the two steps approach previously described, we implicitly removed this indeterminacy by requiring in step 1 that the measure be supported on the boundary (because we seek a function holomorphic throughout the right half-space), and by requiring in step 2 that the measure be discrete in the left half-plane (in fact: a sum of point masses $\sum_1^n a_j \delta_{z_j}$). The discreteness assumption also prevails in 3-D inverse source problems, see Section 4.3 . Conditions that ensure uniqueness of the solution to the inverse potential problem are part of the so-called regularizing assumptions which are needed in each case to derive efficient algorithms.

To recap, the gist of our approach is to approximate boundary data by (boundary traces of) fields arising from potentials of measures with specific support. This differs from standard approaches to inverse problems, where descent algorithms are applied to integration schemes of the direct problem; in such methods, it is the equation which gets approximated (in fact: discretized).

Along these lines, Apics advocates the use of steps 1 and 2 above, along with some singularity analysis, to approach issues of nondestructive control in 2-D and 3-D [2], [5], [40]. The team is currently engaged in the generalization to inverse source problems for the Laplace equation in 3-D, to be described further in Section 3.2.1 . There, holomorphic functions are replaced by harmonic gradients; applications are to EEG/MEG and inverse magnetization problems in geosciences, see Section 4.3 .

The approximation-theoretic tools developed by Apics to handle issues mentioned so far are outlined in Section 3.3 . In Section 3.2 to come, we describe in more detail which problems are considered and which applications are targeted.

3.2. Range of inverse problems

3.2.1. Elliptic partial differential equations (PDE)

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Konstantinos Mavreas, Christos Papageorgakis, Dmitry Ponomarev.

By standard properties of conjugate differentials, reconstructing Dirichlet-Neumann boundary conditions for a function harmonic in a plane domain, when these conditions are already known on a subset E of the boundary, is equivalent to recover a holomorphic function in the domain from its boundary values on E . This is the problem raised on the half-plane in step 1 of Section 3.1 . It makes good sense in holomorphic Hardy spaces where functions are entirely determined by their values on boundary subsets of positive linear measure, which is the framework for Problem (P) that we set up in Section 3.3.1 . Such issues naturally arise in nondestructive testing of 2-D (or 3-D cylindrical) materials from partial electrical measurements on the boundary. For instance, the ratio between the tangential and the normal currents (the so-called Robin coefficient) tells one about corrosion of the material. Thus, solving Problem (P) where ψ is chosen to be the response of some uncorroded piece with identical shape yields non destructive testing of a potentially

⁰There is a subtle difference here between dimension 2 and higher. Indeed, a function holomorphic on a plane domain is defined by its non-tangential limit on a boundary subset of positive linear measure, but there are non-constant harmonic functions in the 3-D ball, C^1 up to the boundary sphere, yet having vanishing gradient on a subset of positive measure of the sphere. Such a "bad" subset, however, cannot have interior points on the sphere.

corroded piece of material, part of which is inaccessible to measurements. This was an initial application of holomorphic extremal problems to non-destructive control [54], [57].

Another application by the team deals with non-constant conductivity over a doubly connected domain, the set E being now the outer boundary. Measuring Dirichlet-Neumann data on E , one wants to recover level lines of the solution to a conductivity equation, which is a so-called free boundary inverse problem. For this, given a closed curve inside the domain, we first quantify how constant the solution on this curve. To this effect, we state and solve an analog of Problem (P), where the constraint bears on the real part of the function on the curve (it should be close to a constant there), in a Hardy space of a conjugate Beltrami equation, of which the considered conductivity equation is the compatibility condition (just like the Laplace equation is the compatibility condition of the Cauchy-Riemann system). Subsequently, a descent algorithm on the curve leads one to improve the initial guess. For example, when the domain is regarded as separating the edge of a tokamak's vessel from the plasma (rotational symmetry makes this a 2-D situation), this method can be used to estimate the shape of a plasma subject to magnetic confinement. This was actually carried out in collaboration with CEA (French nuclear agency) and the University of Nice (JAD Lab.), to data from *Tore Supra* [60]. The procedure is fast because no numerical integration of the underlying PDE is needed, as an explicit basis of solutions to the conjugate Beltrami equation in terms of Bessel functions was found in this case. Generalizing this approach in a more systematic manner to free boundary problems of Bernoulli type, using descent algorithms based on shape-gradient for such approximation-theoretic criteria, is an interesting prospect now under study in the team.

The piece of work we just mentioned requires defining and studying Hardy spaces of the conjugate-Beltrami equation, which is an interesting topic by itself. For Sobolev-smooth coefficients of exponent greater than 2, they were investigated in [4], [34]. The case of the critical exponent 2 is treated in [12], which apparently provides the first example of well-posedness for the Dirichlet problem in the non-strictly elliptic case: the conductivity may be unbounded or zero on sets of zero capacity and, accordingly, solutions need not be locally bounded. Exponent 2 seems also to be the key to a similar theory on general (rectifiable) domains in the plane, for exponent 2 is all one is left with in general after a conformal transformation of the domain.

Generalized Hardy classes as above are used in [13] where we address the uniqueness issue in the classical Robin inverse problem on a Lipschitz domain of $\Omega \subset \mathbb{R}^n$, $n \geq 2$, with uniformly bounded Robin coefficient, L^2 Neumann data and conductivity of Sobolev class $W^{1,r}(\Omega)$, $r > n$. We show that uniqueness of the Robin coefficient on a subset of the boundary, given Cauchy data on the complementary part, does hold in dimension $n = 2$, thanks to a unique continuation result, but needs not hold in higher dimension. In higher dimension, this raises an open issue on harmonic gradients, namely whether the positivity of the Robin coefficient is compatible with identical vanishing of the boundary gradient on a subset of positive measure.

The 3-D version of step 1 in Section 3.1 is another subject investigated by Apics: to recover a harmonic function (up to an additive constant) in a ball or a half-space from partial knowledge of its gradient. This prototypical inverse problem (*i.e.* inverse to the Cauchy problem for the Laplace equation) often recurs in electromagnetism. At present, Apics is involved with solving instances of this inverse problem arising in two fields, namely medical imaging *e.g.* for electroencephalography (EEG) or magneto-encephalography (MEG), and paleomagnetism (recovery of rocks magnetization) [2], [36], see Section 5.1. In this connection, we collaborate with two groups of partners: Athena Inria project-team, CHU La Timone, and BESA company on the one hand, Geosciences Lab. at MIT and Cerege CNRS Lab. on the other hand. The question is considerably more difficult than its 2-D counterpart, due mainly to the lack of multiplicative structure for harmonic gradients. Still, substantial progress has been made over the last years using methods of harmonic analysis and operator theory.

The team is further concerned with 3-D generalizations and applications to non-destructive control of step 2 in Section 3.1. A typical problem is here to localize inhomogeneities or defaults such as cracks, sources or occlusions in a planar or 3-dimensional object, knowing thermal, electrical, or magnetic measurements on the boundary. These defaults can be expressed as a lack of harmonicity of the solution to the associated Dirichlet-Neumann problem, thereby posing an inverse potential problem in order to recover them. In 2-D, finding an optimal discretization of the potential in Sobolev norm amounts to solve a best rational approximation

problem, and the question arises as to how the location of the singularities of the approximant (*i.e.* its poles) reflects the location of the singularities of the potential (*i.e.* the defaults we seek). This is a fairly deep issue in approximation theory, to which Apics contributed convergence results for certain classes of fields expressed as Cauchy integrals over extremal contours for the logarithmic potential [6], [37], [51]. Initial schemes to locate cracks or sources *via* rational approximation on planar domains were obtained this way [40], [44], [54]. It is remarkable that finite inverse source problems in 3-D balls, or more general algebraic surfaces, can be approached using these 2-D techniques upon slicing the domain into planar sections [7], [41]. More precisely, each section cuts out a planar domain, the boundary of which carries data which can be proved to match an algebraic function. The singularities of this algebraic function are not located at the 3-D sources, but are related to them: the section contains a source if and only if some function of the singularities in that section meets a relative extremum. Using bisection it is thus possible to determine an extremal place along all sections parallel to a given plane direction, up to some threshold which has to be chosen small enough that one does not miss a source. This way, we reduce the original source problem in 3-D to a sequence of inverse poles and branchpoints problems in 2-D. This bottom line generates a steady research activity within Apics, and again applications are sought to medical imaging and geosciences, see Sections 4.3 , 4.2 and 5.1 .

Conjectures may be raised on the behavior of optimal potential discretization in 3-D, but answering them is an ambitious program still in its infancy.

3.2.2. Systems, transfer and scattering

Participants: Laurent Baratchart, Matthias Caenepeel, Sylvain Chevillard, Martine Olivi, Fabien Seyfert.

Through contacts with CNES (French space agency), members of the team became involved in identification and tuning of microwave electromagnetic filters used in space telecommunications, see Section 4.4 . The initial problem was to recover, from band-limited frequency measurements, physical parameters of the device under examination. The latter consists of interconnected dual-mode resonant cavities with negligible loss, hence its scattering matrix is modeled by a 2×2 unitary-valued matrix function on the frequency line, say the imaginary axis to fix ideas. In the bandwidth around the resonant frequency, a modal approximation of the Helmholtz equation in the cavities shows that this matrix is approximately rational, of Mc-Millan degree twice the number of cavities.

This is where system theory comes into play, through the so-called *realization* process mapping a rational transfer function in the frequency domain to a state-space representation of the underlying system of linear differential equations in the time domain. Specifically, realizing the scattering matrix allows one to construct a virtual electrical network, equivalent to the filter, the parameters of which mediate in between the frequency response and the geometric characteristics of the cavities (*i.e.* the tuning parameters).

Hardy spaces provide a framework to transform this ill-posed issue into a series of regularized analytic and meromorphic approximation problems. More precisely, the procedure sketched in Section 3.1 goes as follows:

1. infer from the pointwise boundary data in the bandwidth a stable transfer function (*i.e.* one which is holomorphic in the right half-plane), that may be infinite dimensional (numerically: of high degree). This is done by solving a problem analogous to (P) in Section 3.3.1 , while taking into account prior knowledge on the decay of the response outside the bandwidth, see [9] for details.
2. A stable rational approximation of appropriate degree to the model obtained in the previous step is performed. For this, a descent method on the compact manifold of inner matrices of given size and degree is used, based on an original parametrization of stable transfer functions developed within the team [28], [9].
3. Realizations of this rational approximant are computed. To be useful, they must satisfy certain constraints imposed by the geometry of the device. These constraints typically come from the coupling topology of the equivalent electrical network used to model the filter. This network is composed of resonators, coupled according to some specific graph. This realization step can be recast, under appropriate compatibility conditions [55], as solving a zero-dimensional multivariate polynomial system. To tackle this problem in practice, we use Gröbner basis techniques and continuation methods which team up in the Dedale-HF software (see Section 3.4.1).

Let us mention that extensions of classical coupling matrix theory to frequency-dependent (reactive) couplings have been carried-out in recent years [1] for wide-band design applications.

Apics also investigates issues pertaining to design rather than identification. Given the topology of the filter, a basic problem in this connection is to find the optimal response subject to specifications that bear on rejection, transmission and group delay of the scattering parameters. Generalizing the classical approach based on Chebyshev polynomials for single band filters, we recast the problem of multi-band response synthesis as a generalization of the classical Zolotarev min-max problem for rational functions [27] [8]. Thanks to quasi-convexity, the latter can be solved efficiently using iterative methods relying on linear programming. These were implemented in the software easy-FF (see [easy-FF](#)). Currently, the team is engaged in the synthesis of more complex microwave devices like multiplexers and routers, which connect several filters through wave guides. Schur analysis plays an important role here, because scattering matrices of passive systems are of Schur type (*i.e.* contractive in the stability region). The theory originates with the work of I. Schur [75], who devised a recursive test to check for contractivity of a holomorphic function in the disk. The so-called Schur parameters of a function may be viewed as Taylor coefficients for the hyperbolic metric of the disk, and the fact that Schur functions are contractions for that metric lies at the root of Schur's test. Generalizations thereof turn out to be efficient to parametrize solutions to contractive interpolation problems [29]. Dwelling on this, Apics contributed differential parametrizations (atlases of charts) of lossless matrix functions [28], [71], [65] which are fundamental to our rational approximation software RARL2 (see Section 3.4.4). Schur analysis is also instrumental to approach de-embedding issues, and provides one with considerable insight into the so-called matching problem. The latter consists in maximizing the power a multiport can pass to a given load, and for reasons of efficiency it is all-pervasive in microwave and electric network design, *e.g.* of antennas, multiplexers, wifi cards and more. It can be viewed as a rational approximation problem in the hyperbolic metric, and the team presently deals with this hot topic using contractive interpolation with constraints on boundary peak points, within the framework of the (defense funded) ANR COCORAM, see Sections 5.2 and 7.2.1.

In recent years, our attention was driven by CNES and UPV (Bilbao) to questions about stability of high-frequency amplifiers, see Section 6.2. Contrary to previously discussed devices, these are *active* components. The response of an amplifier can be linearized around a set of primary current and voltages, and then admittances of the corresponding electrical network can be computed at various frequencies, using the so-called harmonic balance method. The initial goal is to check for stability of the linearized model, so as to ascertain existence of a well-defined working state. The network is composed of lumped electrical elements namely inductors, capacitors, negative *and* positive reactors, transmission lines, and controlled current sources. Our research so far has focused on describing the algebraic structure of admittance functions, so as to set up a function-theoretic framework where the two-steps approach outlined in Section 3.1 can be put to work. The main discovery is that the unstable part of each partial transfer function is rational and can be computed by analytic projection, see Section 5.4. We now start investigating the linearized harmonic transfer-function around a periodic cycle, to check for stability under non necessarily small inputs. This generalization generates both doctoral and postdoctoral work by new students in the team.

3.3. Approximation

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Martine Olivi, Dmitry Ponomarev, Fabien Seyfert.

3.3.1. Best analytic approximation

In dimension 2, the prototypical problem to be solved in step 1 of Section 3.1 may be described as: given a domain $D \subset \mathbb{R}^2$, to recover a holomorphic function from its values on a subset K of the boundary of D . For the discussion it is convenient to normalize D , which can be done by conformal mapping. So, in the simply connected case, we fix D to be the unit disk with boundary unit circle T . We denote by H^p the Hardy space of exponent p , which is the closure of polynomials in $L^p(T)$ -norm if $1 \leq p < \infty$ and the space of bounded holomorphic functions in D if $p = \infty$. Functions in H^p have well-defined boundary values in $L^p(T)$, which makes it possible to speak of (traces of) analytic functions on the boundary.

To find an analytic function g in D matching some measured values f approximately on a sub-arc K of T , we formulate a constrained best approximation problem as follows.

(P) Let $1 \leq p \leq \infty$, K a sub-arc of T , $f \in L^p(K)$, $\psi \in L^p(T \setminus K)$ and $M > 0$; find a function $g \in H^p$ such that $\|g - \psi\|_{L^p(T \setminus K)} \leq M$ and $g - f$ is of minimal norm in $L^p(K)$ under this constraint.

Here ψ is a reference behavior capturing *a priori* assumptions on the behavior of the model off K , while M is some admissible deviation thereof. The value of p reflects the type of stability which is sought and how much one wants to smooth out the data. The choice of L^p classes is suited to handle pointwise measurements.

To fix terminology, we refer to (P) as a *bounded extremal problem*. As shown in [39], [42], [48], the solution to this convex infinite-dimensional optimization problem can be obtained when $p \neq 1$ upon iterating with respect to a Lagrange parameter the solution to spectral equations for appropriate Hankel and Toeplitz operators. These spectral equations involve the solution to the special case $K = T$ of (P), which is a standard extremal problem [63]:

(P_0) Let $1 \leq p \leq \infty$ and $\varphi \in L^p(T)$; find a function $g \in H^p$ such that $g - \varphi$ is of minimal norm in $L^p(T)$.

The case $p = 1$ is more or less open.

Various modifications of (P) can be tailored to meet specific needs. For instance when dealing with lossless transfer functions (see Section 4.4), one may want to express the constraint on $T \setminus K$ in a pointwise manner: $|g - \psi| \leq M$ a.e. on $T \setminus K$, see [43]. In this form, the problem comes close to (but still is different from) H^∞ frequency optimization used in control [66], [74]. One can also impose bounds on the real or imaginary part of $g - \psi$ on $T \setminus K$, which is useful when considering Dirichlet-Neumann problems, see [68].

The analog of Problem (P) on an annulus, K being now the outer boundary, can be seen as a means to regularize a classical inverse problem occurring in nondestructive control, namely to recover a harmonic function on the inner boundary from Dirichlet-Neumann data on the outer boundary (see Sections 3.2.1, 4.3, 5.1.5). It may serve as a tool to approach Bernoulli type problems, where we are given data on the outer boundary and we *seek the inner boundary*, knowing it is a level curve of the solution. In this case, the Lagrange parameter indicates how to deform the inner contour in order to improve data fitting. Similar topics are discussed in Section 3.2.1 for more general equations than the Laplacian, namely isotropic conductivity equations of the form $\operatorname{div}(\sigma \nabla u) = 0$ where σ is no longer constant. Then, the Hardy spaces in Problem (P) are those of a so-called conjugate Beltrami equation: $\bar{\partial} f = \nu \bar{\partial} f$ [67], which are studied for $1 < p < \infty$ in [4], [12], [34] and [58]. Expansions of solutions needed to constructively handle such issues in the specific case of linear fractional conductivities (occurring for instance in plasma shaping) have been expounded in [60].

Another instance of problem (P) in with $p = 2$ and additional pointwise interpolation constraints inside a simply connected domain (disk) D was studied and solved in [11], Part I, and [15]. Such pointwise interpolation constraints could be of practical interest for inverse Cauchy type problems in cases where interior information is also available or to model uncertainty on boundary data.

Though originally considered in dimension 2, Problem (P) carries over naturally to higher dimensions where analytic functions get replaced by gradients of harmonic functions. Namely, given some open set $\Omega \subset \mathbb{R}^n$ and some \mathbb{R}^n -valued vector field V on an open subset O of the boundary of Ω , we seek a harmonic function in Ω whose gradient is close to V on O .

When Ω is a ball or a half-space, a substitute for holomorphic Hardy spaces is provided by the Stein-Weiss Hardy spaces of harmonic gradients [77]. Conformal maps are no longer available when $n > 2$, so that Ω can no longer be normalized. More general geometries than spheres and half-spaces have not been much studied so far.

On the ball, the analog of Problem (P) is

(P_1) Let $1 \leq p \leq \infty$ and $B \subset \mathbb{R}^n$ the unit ball. Fix O an open subset of the unit sphere $S \subset \mathbb{R}^n$. Let further $V \in L^p(O)$ and $W \in L^p(S \setminus O)$ be \mathbb{R}^n -valued vector fields. Given $M > 0$, find a

harmonic gradient $G \in H^p(B)$ such that $\|G - W\|_{L^p(S \setminus O)} \leq M$ and $G - V$ is of minimal norm in $L^p(O)$ under this constraint.

When $p = 2$, Problem (P_1) was solved in [2] as well as its analog on a shell, when the tangential component of V is a gradient (when O is Lipschitz the general case follows easily from this). The solution extends the work in [39] to the 3-D case, using a generalization of Toeplitz operators. The case of the shell was motivated by applications to the processing of EEG data. An important ingredient is a refinement of the Hodge decomposition, that we call the *Hardy-Hodge* decomposition, allowing us to express a \mathbb{R}^n -valued vector field in $L^p(S)$, $1 < p < \infty$, as the sum of a vector field in $H^p(B)$, a vector field in $H^p(\mathbb{R}^n \setminus \overline{B})$, and a tangential divergence free vector field on S ; the space of such divergence-free fields is denoted by $D(S)$. If $p = 1$ or $p = \infty$, L^p must be replaced by the real Hardy space or the space of functions with bounded mean oscillation. More generally this decomposition, which is valid on any sufficiently smooth surface (see Section 5.1), seems to play a fundamental role in inverse potential problems. In fact, it was first introduced formally on the plane to describe silent magnetizations supported in \mathbb{R}^2 (i.e. those generating no field in the upper half space) [36].

Just like solving problem (P) appeals to the solution of problem (P_0) , our ability to solve problem (P_1) will depend on the possibility to tackle the special case where $O = S$:

(P_2) Let $1 \leq p \leq \infty$ and $V \in L^p(S)$ be a \mathbb{R}^n -valued vector field. Find a harmonic gradient $G \in H^p(B)$ such that $\|G - V\|_{L^p(S)}$ is minimum.

Problem (P_2) is simple when $p = 2$ by virtue of the Hardy Hodge decomposition together with orthogonality of $H^2(B)$ and $H^2(\mathbb{R}^n \setminus \overline{B})$, which is the reason why we were able to solve (P_1) in this case. Other values of p cannot be treated as easily and are still under investigation, especially the case $p = \infty$ which is of particular interest and presents itself as a 3-D analog to the Nehari problem [73].

Companion to problem (P_2) is problem (P_3) below.

(P_3) Let $1 \leq p \leq \infty$ and $V \in L^p(S)$ be a \mathbb{R}^n -valued vector field. Find $G \in H^p(B)$ and $D \in D(S)$ such that $\|G + D - V\|_{L^p(S)}$ is minimum.

Note that (P_2) and (P_3) are identical in 2-D, since no non-constant tangential divergence-free vector field exists on T . It is no longer so in higher dimension, where both (P_2) and (P_3) arise in connection with inverse potential problems in divergence form, like source recovery in electro/magneto encephalography and paleomagnetism, see Sections 3.2.1 and 4.3.

3.3.2. Best meromorphic and rational approximation

The techniques set forth in this section are used to solve step 2 in Section 3.2 and they are instrumental to approach inverse boundary value problems for the Poisson equation $\Delta u = \mu$, where μ is some (unknown) measure.

3.3.2.1. Scalar meromorphic and rational approximation

We put R_N for the set of rational functions with at most N poles in D . By definition, meromorphic functions in $L^p(T)$ are (traces of) functions in $H^p + R_N$.

A natural generalization of problem (P_0) is:

(P_N) Let $1 \leq p \leq \infty$, $N \geq 0$ an integer, and $f \in L^p(T)$; find a function $g_N \in H^p + R_N$ such that $g_N - f$ is of minimal norm in $L^p(T)$.

Only for $p = \infty$ and f continuous is it known how to solve (P_N) in semi-closed form. The unique solution is given by AAK theory (named after Adamjan, Arov and Krein), which connects the spectral decomposition of Hankel operators with best approximation [73].

The case where $p = 2$ is of special importance for it reduces to rational approximation. Indeed, if we write the Hardy decomposition $f = f^+ + f^-$ where $f^+ \in H^2$ and $f^- \in H^2(\mathbb{C} \setminus \overline{D})$, then $g_N = f^+ + r_N$ where r_N is a best approximant to f^- from R_N in $L^2(T)$. Moreover, r_N has no pole outside D , hence it is a *stable* rational approximant to f^- . However, in contrast to the case where $p = \infty$, this best approximant may *not* be unique.

The former Miaou project (predecessor of Apics) designed a dedicated steepest-descent algorithm for the case $p = 2$ whose convergence to a *local minimum* is guaranteed; until now it seems to be the only procedure meeting this property. This gradient algorithm proceeds recursively with respect to N on a compactification of the parameter space [32]. Although it has proved to be effective in all applications carried out so far (see Sections 4.3 , 4.4), it is still unknown whether the absolute minimum can always be obtained by choosing initial conditions corresponding to *critical points* of lower degree (as is done by the RARL2 software, Section 3.4.4).

In order to establish global convergence results, Apics has undertaken a deeper study of the number and nature of critical points (local minima, saddle points...), in which tools from differential topology and operator theory team up with classical interpolation theory [45], [47]. Based on this work, uniqueness or asymptotic uniqueness of the approximant was proved for certain classes of functions like transfer functions of relaxation systems (*i.e.* Markov functions) [49] and more generally Cauchy integrals over hyperbolic geodesic arcs [52]. These are the only results of this kind. Research by Apics on this topic remained dormant for a while by reasons of opportunity, but revisiting the work [30] in higher dimension is a worthy and timely endeavor today. Meanwhile, an analog to AAK theory was carried out for $2 \leq p < \infty$ in [48]. Although not as effective computationally, it was recently used to derive lower bounds [3]. When $1 \leq p < 2$, problem (P_N) is still quite open.

A common feature to the above-mentioned problems is that critical point equations yield non-Hermitian orthogonality relations for the denominator of the approximant. This stresses connections with interpolation, which is a standard way to build approximants, and in many respects best or near-best rational approximation may be regarded as a clever manner to pick interpolation points. This was exploited in [53], [50], and is used in an essential manner to assess the behavior of poles of best approximants to functions with branched singularities, which is of particular interest for inverse source problems (*cf.* Sections 3.4.2 and 5.1).

In higher dimensions, the analog of Problem (P_N) is best approximation of a vector field by gradients of discrete potentials generated by N point masses. This basic issue is by no means fully understood, and it is an exciting field of research. It is connected with certain generalizations of Toeplitz or Hankel operators, and with constructive approaches to so-called weak factorizations for real Hardy functions [59].

Besides, certain constrained rational approximation problems, of special interest in identification and design of passive systems, arise when putting additional requirements on the approximant, for instance that it should be smaller than 1 in modulus (*i.e.* a Schur function). In particular, Schur interpolation lately received renewed attention from the team, in connection with matching problems. There, interpolation data are subject to a well-known compatibility condition (positive definiteness of the so-called Pick matrix), and the main difficulty is to put interpolation points on the boundary of D while controlling both the degree and the extremal points (peak points for the modulus) of the interpolant. Results obtained by Apics in this direction generalize a variant of contractive interpolation with degree constraint as studied in [64], see Section 5.2 . We mention that contractive interpolation with nodes approaching the boundary has been a subsidiary research topic by the team in the past, which plays an interesting role in the spectral representation of certain non-stationary stochastic processes [35], [38]. The subject is intimately connected to orthogonal polynomials on the unit circle, and this line of investigation has been pursued towards an asymptotic study of orthogonal polynomials on planar domains, which is today an active area in approximation theory with application to quantum particle systems, spectra of random matrices, and Hele-Shaw flows, see Section 5.6 .

3.3.2.2. Matrix-valued rational approximation

Matrix-valued approximation is necessary to handle systems with several inputs and outputs but it generates additional difficulties as compared to scalar-valued approximation, both theoretically and algorithmically. In the matrix case, the McMillan degree (*i.e.* the degree of a minimal realization in the System-Theoretic sense) generalizes the usual notion of degree for rational functions. For instance when poles are simple, the McMillan degree is the sum of the ranks of the residues.

The basic problem that we consider now goes as follows: let $\mathcal{F} \in (H^2)^{m \times l}$ and n an integer; find a rational matrix of size $m \times l$ without poles in the unit disk and of McMillan degree at most n which is nearest possible

to \mathcal{F} in $(H^2)^{m \times l}$. Here the L^2 norm of a matrix is the square root of the sum of the squares of the norms of its entries.

The scalar approximation algorithm derived in [32] and mentioned in Section 3.3.2.1 generalizes to the matrix-valued situation [62]. The first difficulty here is to parametrize inner matrices (*i.e.* matrix-valued functions analytic in the unit disk and unitary on the unit circle) of given McMillan degree n . Indeed, inner matrices play the role of denominators in fractional representations of transfer matrices (using the so-called Douglas-Shapiro-Shields factorization). The set of inner matrices of given degree is a smooth manifold that allows one to use differential tools as in the scalar case. In practice, one has to produce an atlas of charts (local parametrizations) and to handle changes of charts in the course of the algorithm. Such parametrization can be obtained using interpolation theory and Schur-type algorithms, the parameters of which are vectors or matrices ([28], [65], [71]). Some of these parametrizations are also interesting to compute realizations and achieve filter synthesis ([65], [71]). The rational approximation software “RARL2” developed by the team is described in Section 3.4.4.

Difficulties relative to multiple local minima of course arise in the matrix-valued case as well, and deriving criteria that guarantee uniqueness is even more difficult than in the scalar case. The case of rational functions of degree n or small perturbations thereof (the consistency problem) was solved in [46]. Matrix-valued Markov functions are the only known example beyond this one [31].

Let us stress that RARL2 seems the only algorithm handling rational approximation in the matrix case that demonstrably converges to a local minimum while meeting stability constraints on the approximant. It is still a working pin of many developments by Apics on frequency optimization and design.

3.3.3. Behavior of poles of meromorphic approximants

Participant: Laurent Baratchart.

We refer here to the behavior of poles of best meromorphic approximants, in the L^p -sense on a closed curve, to functions f defined as Cauchy integrals of complex measures whose support lies inside the curve. Normalizing the contour to be the unit circle T , we are back to Problem (P_N) in Section 3.3.2.1 ; invariance of the latter under conformal mapping was established in [5]. Research so far has focused on functions whose singular set inside the contour is polar, meaning that the function can be continued analytically (possibly in a multiple-valued manner) except over a set of logarithmic capacity zero.

Generally speaking in approximation theory, assessing the behavior of poles of rational approximants is essential to obtain error rates as the degree goes large, and to tackle constructive issues like uniqueness. However, as explained in Section 3.2.1, the original twist by Apics is to consider this issue also as a means to extract information on singularities of the solution to a Dirichlet-Neumann problem. The general theme is thus: *how do the singularities of the approximant reflect those of the approximated function?* This approach to inverse problem for the 2-D Laplacian turns out to be attractive when singularities are zero- or one-dimensional (see Section 4.3). It can be used as a computationally cheap initial condition for more precise but much heavier numerical optimizations which often do not even converge unless properly initialized. As regards crack detection or source recovery, this approach boils down to analyzing the behavior of best meromorphic approximants of given pole cardinality to a function with branch points, which is the prototype of a polar singular set. For piecewise analytic cracks, or in the case of sources, we were able to prove ([5], [6], [37]), that the poles of the approximants accumulate, when the degree goes large, to some extremal cut of minimum weighted logarithmic capacity connecting the singular points of the crack, or the sources [40]. Moreover, the asymptotic density of the poles turns out to be the Green equilibrium distribution on this cut in D , therefore it charges the singular points if one is able to approximate in sufficiently high degree (this is where the method could fail, because high-order approximation requires rather precise data).

The case of two-dimensional singularities is still an outstanding open problem.

It is remarkable that inverse source problems inside a sphere or an ellipsoid in 3-D can be approached with such 2-D techniques, as applied to planar sections, see Section 5.1. The technique is implemented in the software FindSources3D, see Section 3.4.2.

3.4. Software tools of the team

In addition to the above-mentioned research activities, Apics develops and maintains a number of long-term software tools that either implement and illustrate effectiveness of the algorithms theoretically developed by the team or serve as tools to help further research by team members. We present briefly the most important of them.

3.4.1. DEDALE-HF

SCIENTIFIC DESCRIPTION

Dedale-HF consists in two parts: a database of coupling topologies as well as a dedicated predictor-corrector code. Roughly speaking each reference file of the database contains, for a given coupling topology, the complete solution to the coupling matrix synthesis problem (C.M. problem for short) associated to particular filtering characteristics. The latter is then used as a starting point for a predictor-corrector integration method that computes the solution to the C.M. corresponding to the user-specified filter characteristics. The reference files are computed off-line using Gröbner basis techniques or numerical techniques based on the exploration of a monodromy group. The use of such continuation techniques, combined with an efficient implementation of the integrator, drastically reduces the computational time.

Dedale-HF has been licensed to, and is currently used by TAS-Espana

FUNCTIONAL DESCRIPTION

Dedale-HF is a software dedicated to solve exhaustively the coupling matrix synthesis problem in reasonable time for the filtering community. Given a coupling topology, the coupling matrix synthesis problem consists in finding all possible electromagnetic coupling values between resonators that yield a realization of given filter characteristics. Solving the latter is crucial during the design step of a filter in order to derive its physical dimensions, as well as during the tuning process where coupling values need to be extracted from frequency measurements.

- Participant: Fabien Seyfert
- Contact: Fabien Seyfert
- URL: <http://www-sop.inria.fr/apics/Dedale/>

3.4.2. FindSources3D

FindSources3D-bolis

KEYWORDS: Health - Neuroimaging - Visualization - Compilers - Medical - Image - Processing

FUNCTIONAL DESCRIPTION

FindSources3D is a software program dedicated to the resolution of inverse source problems in electroencephalography (EEG). From pointwise measurements of the electrical potential taken by electrodes on the scalp, FindSources3D estimates pointwise dipolar current sources within the brain in a spherical model.

After a first data transmission “cortical mapping” step, it makes use of best rational approximation on 2-D planar cross-sections and of the software RARL2 in order to locate singularities. From those planar singularities, the 3-D sources are estimated in a last step.

This version of FindSources3D provides a modular, ergonomic, accessible and interactive platform, with a convenient graphical interface and a tool that can be distributed and used, for EEG medical imaging. Modularity is now granted (using the tools dtk, Qt, with compiled Matlab libraries). It offers a detailed and nice visualization of data and tuning parameters, processing steps, and of the computed results (using VTK).

- Participants: Juliette Leblond, Maureen Clerc Gallagher, Théodore Papadopoulo, Jean-Paul Marmorat and Nicolas Schnitzler
- Contact: Juliette Leblond
- URL: <http://www-sop.inria.fr/apics/FindSources3D/en/index.html>

3.4.3. PRESTO-HF

SCIENTIFIC DESCRIPTION

For the matrix-valued rational approximation step, Presto-HF relies on RARL2. Constrained realizations are computed using the Dedale-HF software. As a toolbox, Presto-HF has a modular structure, which allows one for example to include some building blocks in an already existing software.

The delay compensation algorithm is based on the following assumption: far off the pass-band, one can reasonably expect a good approximation of the rational components of S11 and S22 by the first few terms of their Taylor expansion at infinity, a small degree polynomial in $1/s$. Using this idea, a sequence of quadratic convex optimization problems are solved, in order to obtain appropriate compensations. In order to check the previous assumption, one has to measure the filter on a larger band, typically three times the pass band.

This toolbox has been licensed to, and is currently used by Thales Alenia Space in Toulouse and Madrid, Thales airborne systems and Flextronics (two licenses). XLIM (University of Limoges) is a heavy user of Presto-HF among the academic filtering community and some free license agreements have been granted to the microwave department of the University of Erlangen (Germany) and the Royal Military College (Kingston, Canada).

FUNCTIONAL DESCRIPTION

Presto-HF is a toolbox dedicated to low-pass parameter identification for microwave filters. In order to allow the industrial transfer of our methods, a Matlab-based toolbox has been developed, dedicated to the problem of identification of low-pass microwave filter parameters. It allows one to run the following algorithmic steps, either individually or in a single stroke:

- Determination of delay components caused by the access devices (automatic reference plane adjustment),
- Automatic determination of an analytic completion, bounded in modulus for each channel,
- Rational approximation of fixed McMillan degree,
- Determination of a constrained realization.
 - Participants: Fabien Seyfert, Jean-Paul Marmorat and Martine Olivi
 - Contact: Fabien Seyfert
 - URL: <https://project.inria.fr/presto-hf/>

3.4.4. RARL2

Réalisation interne et Approximation Rationnelle L2

SCIENTIFIC DESCRIPTION

The method is a steepest-descent algorithm. A parametrization of MIMO systems is used, which ensures that the stability constraint on the approximant is met. The implementation, in Matlab, is based on state-space representations.

RARL2 performs the rational approximation step in the software tools PRESTO-HF and FindSources3D. It is distributed under a particular license, allowing unlimited usage for academic research purposes. It was released to the universities of Delft and Maastricht (the Netherlands), Cork (Ireland), Brussels (Belgium), Macao (China) and BITS-Pilani Hyderabad Campus (India).

FUNCTIONAL DESCRIPTION

RARL2 is a software for rational approximation. It computes a stable rational L2-approximation of specified order to a given L2-stable (L2 on the unit circle, analytic in the complement of the unit disk) matrix-valued function. This can be the transfer function of a multivariable discrete-time stable system. RARL2 takes as input either:

- its internal realization,
- its first N Fourier coefficients,

- discretized (uniformly distributed) values on the circle. In this case, a least-square criterion is used instead of the L2 norm.

It thus performs model reduction in the first or the second case, and leans on frequency data identification in the third. For band-limited frequency data, it could be necessary to infer the behavior of the system outside the bandwidth before performing rational approximation.

An appropriate Möbius transformation allows to use the software for continuous-time systems as well.

- Participants: Jean-Paul Marmorat and Martine Olivi
- Contact: Martine Olivi
- URL: <http://www-sop.inria.fr/apics/RARL2/rar12.html>

3.4.5. Sollya

KEYWORDS: Numerical algorithm - Supremum norm - Curve plotting - Remez algorithm - Code generator - Proof synthesis

FUNCTIONAL DESCRIPTION

Sollya is an interactive tool where the developers of mathematical floating-point libraries (libm) can experiment before actually developing code. The environment is safe with respect to floating-point errors, i.e. the user precisely knows when rounding errors or approximation errors happen, and rigorous bounds are always provided for these errors.

Among other features, it offers a fast Remez algorithm for computing polynomial approximations of real functions and also an algorithm for finding good polynomial approximants with floating-point coefficients to any real function. As well, it provides algorithms for the certification of numerical codes, such as Taylor Models, interval arithmetic or certified supremum norms.

It is available as a free software under the CeCILL-C license.

- Participants: Sylvain Chevillard, Christoph Lauter, Mioara Joldes and Nicolas Jourdan
- Partners: CNRS - ENS Lyon - UCBL Lyon 1
- Contact: Sylvain Chevillard
- URL: <http://sollya.gforge.inria.fr/>

ECUADOR Project-Team

3. Research Program

3.1. Algorithmic Differentiation

Participants: Laurent Hascoët, Valérie Pascual, Ala Taftaf.

algorithmic differentiation (AD, aka Automatic Differentiation) Transformation of a program, that returns a new program that computes derivatives of the initial program, i.e. some combination of the partial derivatives of the program's outputs with respect to its inputs.

adjoint Mathematical manipulation of the Partial Differential Equations that define a problem, obtaining new differential equations that define the gradient of the original problem's solution.

checkpointing General trade-off technique, used in adjoint AD, that trades duplicate execution of a part of the program to save some memory space that was used to save intermediate results.

Algorithmic Differentiation (AD) differentiates *programs*. The input of AD is a source program P that, given some $X \in \mathbb{R}^n$, returns some $Y = F(X) \in \mathbb{R}^m$, for a differentiable F . AD generates a new source program P' that, given X , computes some derivatives of F [6].

Any execution of P amounts to a sequence of instructions, which is identified with a composition of vector functions. Thus, if

$$\begin{aligned} P & \text{ runs } \{I_1; I_2; \dots; I_p\}, \\ F & \text{ then is } f_p \circ f_{p-1} \circ \dots \circ f_1, \end{aligned} \quad (2)$$

where each f_k is the elementary function implemented by instruction I_k . AD applies the chain rule to obtain derivatives of F . Calling X_k the values of all variables after instruction I_k , i.e. $X_0 = X$ and $X_k = f_k(X_{k-1})$, the Jacobian of F is

$$F'(X) = f'_p(X_{p-1}) \cdot f'_{p-1}(X_{p-2}) \cdot \dots \cdot f'_1(X_0) \quad (3)$$

which can be mechanically written as a sequence of instructions I'_k . This can be generalized to higher level derivatives, Taylor series, etc. Combining the I'_k with the control of P yields P' , and therefore this differentiation is piecewise.

In practice, many applications only need cheaper projections of $F'(X)$ such as:

- **Sensitivities**, defined for a given direction \dot{X} in the input space as:

$$F'(X) \cdot \dot{X} = f'_p(X_{p-1}) \cdot f'_{p-1}(X_{p-2}) \cdot \dots \cdot f'_1(X_0) \cdot \dot{X} \quad (4)$$

This expression is easily computed from right to left, interleaved with the original program instructions. This is the *tangent mode* of AD.

- **Adjoints**, defined after transposition (F'^*), for a given weighting \bar{Y} of the outputs as:

$$F'^*(X) \cdot \bar{Y} = f'^*_1(X_0) \cdot f'^*_2(X_1) \cdot \dots \cdot f'^*_{p-1}(X_{p-2}) \cdot f'^*_p(X_{p-1}) \cdot \bar{Y} \quad (5)$$

This expression is most efficiently computed from right to left, because matrix \times vector products are cheaper than matrix \times matrix products. This is the *adjoint mode* of AD, most effective for optimization, data assimilation [37], adjoint problems [32], or inverse problems.

Adjoint AD builds a very efficient program [34], which computes the gradient in a time independent from the number of parameters n . In contrast, computing the same gradient with the *tangent mode* would require running the tangent differentiated program n times.

However, the X_k are required in the *inverse* of their computation order. If the original program *overwrites* a part of X_k , the differentiated program must restore X_k before it is used by $f'_{k+1}^*(X_k)$. Therefore, the central research problem of adjoint AD is to make the X_k available in reverse order at the cheapest cost, using strategies that combine storage, repeated forward computation from available previous values, or even inverted computation from available later values.

Another research issue is to make the AD model cope with the constant evolution of modern language constructs. From the old days of Fortran77, novelties include pointers and dynamic allocation, modularity, structured data types, objects, vectorial notation and parallel programming. We keep developing our models and tools to handle these new constructs.

3.2. Static Analysis and Transformation of programs

Participants: Laurent Hascoët, Valérie Pascual, Ala Taftaf.

abstract syntax tree Tree representation of a computer program, that keeps only the semantically significant information and abstracts away syntactic sugar such as indentation, parentheses, or separators.

control flow graph Representation of a procedure body as a directed graph, whose nodes, known as basic blocks, each contain a sequence of instructions and whose arrows represent all possible control jumps that can occur at run-time.

abstract interpretation Model that describes program static analysis as a special sort of execution, in which all branches of control switches are taken concurrently, and where computed values are replaced by abstract values from a given *semantic domain*. Each particular analysis gives birth to a specific semantic domain.

data flow analysis Program analysis that studies how a given property of variables evolves with execution of the program. Data Flow analysis is static, therefore studying all possible run-time behaviors and making conservative approximations. A typical data-flow analysis is to detect, at any location in the source program, whether a variable is initialized or not.

The most obvious example of a program transformation tool is certainly a compiler. Other examples are program translators, that go from one language or formalism to another, or optimizers, that transform a program to make it run better. AD is just one such transformation. These tools share the technological basis that lets them implement the sophisticated analyses [25] required. In particular there are common mathematical models to specify these analyses and analyze their properties.

An important principle is *abstraction*: the core of a compiler should not bother about syntactic details of the compiled program. The optimization and code generation phases must be independent from the particular input programming language. This is generally achieved using language-specific *front-ends*, language-independent *middle-ends*, and target-specific *back-ends*. In the middle-end, analysis can concentrate on the semantics of a reduced set of constructs. This analysis operates on an abstract representation of programs made of one *call graph*, whose nodes are themselves *flow graphs* whose nodes (*basic blocks*) contain abstract *syntax trees* for the individual atomic instructions. To each level are attached symbol tables, nested to capture scoping.

Static program analysis can be defined on this internal representation, which is largely language independent. The simplest analyses on trees can be specified with inference rules [28], [35], [26]. But many *data-flow analyses* are more complex, and better defined on graphs than on trees. Since both call graphs and flow graphs may be cyclic, these global analyses will be solved iteratively. *Abstract Interpretation* [29] is a theoretical framework to study complexity and termination of these analyses.

Data flow analyses must be carefully designed to avoid or control combinatorial explosion. At the call graph level, they can run bottom-up or top-down, and they yield more accurate results when they take into account the different call sites of each procedure, which is called *context sensitivity*. At the flow graph level, they can run forwards or backwards, and yield more accurate results when they take into account only the possible execution flows resulting from possible control, which is called *flow sensitivity*.

Even then, data flow analyses are limited, because they are static and thus have very little knowledge of actual run-time values. Far before reaching the very theoretical limit of *undecidability*, one reaches practical limitations to how much information one can infer from programs that use arrays [41], [30] or pointers. Therefore, conservative *over-approximations* must be made, leading to derivative code less efficient than ideal.

3.3. Algorithmic Differentiation and Scientific Computing

Participants: Alain Dervieux, Laurent Hascoët, Bruno Koobus.

linearization In Scientific Computing, the mathematical model often consists of Partial Differential Equations, that are discretized and then solved by a computer program. Linearization of these equations, or alternatively linearization of the computer program, predict the behavior of the model when small perturbations are applied. This is useful when the perturbations are effectively small, as in acoustics, or when one wants the sensitivity of the system with respect to one parameter, as in optimization.

adjoint state Consider a system of Partial Differential Equations that define some characteristics of a system with respect to some input parameters. Consider one particular scalar characteristic. Its sensitivity, (or gradient) with respect to the input parameters can be defined as the solution of “adjoint” equations, deduced from the original equations through linearization and transposition. The solution of the adjoint equations is known as the adjoint state.

Scientific Computing provides reliable simulations of complex systems. For example it is possible to *simulate* the steady or unsteady 3D air flow around a plane that captures the physical phenomena of shocks and turbulence. Next comes *optimization*, one degree higher in complexity because it repeatedly simulates and applies gradient-based optimization steps until an optimum is reached. The next sophistication is *robustness* i.e. to detect and to lower preference to a solution which, although maybe optimal, is very sensitive to uncertainty on design parameters or on manufacturing tolerances. This makes second derivative come into play. Similarly *Uncertainty Quantification* can use second derivatives to evaluate how uncertainty on the simulation inputs imply uncertainty on its outputs.

We investigate several approaches to obtain the gradient, between two extremes:

- One can write an *adjoint system* of mathematical equations, then discretize it and program it by hand. This is time consuming. Although this looks mathematically sound [32], this does not provide the gradient of the discretized function itself, thus degrading the final convergence of gradient-descent optimization.
- One can apply adjoint AD (cf 3.1) on the program that discretizes and solves the direct system. This gives exactly the adjoint of the discrete function computed by the program. Theoretical results [31] guarantee convergence of these derivatives when the direct program converges. This approach is highly mechanizable, but leads to massive use of storage and may require code transformation by hand [36], [39] to reduce memory usage.

If for instance the model is steady, or when the computation uses a Fixed-Point iteration, tradeoffs exist between these two extremes [33], [27] that combine low storage consumption with possible automated adjoint generation. We advocate incorporating them into the AD model and into the AD tools.

MCTAO Project-Team

3. Research Program

3.1. Control Systems

Our effort is directed toward efficient methods for the *control* of real (physical) systems, based on a *model* of the system to be controlled. *System* refers to the physical plant or device, whereas *model* refers to a mathematical representation of it.

We mostly investigate nonlinear systems whose nonlinearities admit a strong structure derived from physics; the equations governing their behavior are then well known, and the modeling part consists in choosing what phenomena are to be kept in the model used for control design, the other phenomena being treated as perturbations; a more complete model may be used for simulations, for instance. We focus on systems that admit a reliable finite-dimensional model, in continuous time; this means that models are controlled ordinary differential equations, often nonlinear.

Choosing accurate models yet simple enough to allow control design is in itself a key issue; however, modeling or identification as a theory is not per se in the scope of our project.

The extreme generality and versatility of linear control do not contradict the often heard sentence “most real life systems are nonlinear”. Indeed, for many control problems, a linear model is sufficient to capture the important features for control. The reason is that most control objectives are local, first order variations around an operating point or a trajectory are governed by a linear control model, and except in degenerate situations (non-controllability of this linear model), the local behavior of a nonlinear dynamic phenomenon is dictated by the behavior of first order variations. Linear control is the hard core of control theory and practice; it has been pushed to a high degree of achievement –see for instance some classics: [64], [55]– that lead to big successes in industrial applications (PID, Kalman filtering, frequency domain design, H^∞ robust control, etc...), it is taught to future engineers, and it is still a topic of ongoing research.

Linear control by itself however reaches its limits in some important situations:

1. **Non local control objectives.** Steering the system from a region to a reasonably remote other one, as in path planning and optimal control, is outside the scope of information given by a local linear approximation. It is why these are by essence nonlinear.
Stabilisation with a basin of attraction larger than the region where the linear approximation is dominant also needs more information than one linear approximation.
2. **Local control at degenerate equilibria.** Linear control yields local stabilization of an equilibrium point based on the tangent linear approximation if the latter is controllable. It is *not* the case at interesting operating points of some physical systems; linear control is irrelevant and specific nonlinear techniques have to be designed. This is an extreme case of the second part of the above item: the region where the linear approximation is dominant vanishes.
3. **Small controls.** In some situations, actuators only allow a very small magnitude of the effect of control compared to the effect of other phenomena. Then the behavior of the system without control plays a major role and we are again outside the scope of linear control methods.

3.2. Structure of nonlinear control systems

In most problems, choosing the proper coordinates, or the right quantities that describe a phenomenon, sheds light on a path to the solution. In control systems, it is often crucial to analyze the structure of the model, deduced from physical principles, of the plant to be controlled; this may lead to putting it via some transformations in a simpler form, or a form that is most suitable for control design. For instance, equivalence to a linear system may allow to use linear control; also, the so-called “flatness” property drastically simplifies path planning [59], [70].

A better understanding of the “set of nonlinear models”, partly classifying them, has another motivation than facilitating control design for a given system and its model: it may also be a necessary step towards a theory of “nonlinear identification” and modeling. Linear identification is a mature area of control science; its success is mostly due to a very fine knowledge of the structure of the class of linear models: similarly, any progress in the understanding of the structure of the class of nonlinear models would be a contribution to a possible theory of nonlinear identification.

These topics are central in control theory, but raise very difficult mathematical questions: static feedback classification is a geometric problem which is feasible in principle, although describing invariants explicitly is technically very difficult; and conditions for dynamic feedback equivalence and linearization raise unsolved mathematical problems, that make one wonder about decidability⁰.

3.3. Optimal control and feedback control, stabilization

3.3.1. Optimal control.

Mathematically speaking, optimal control is the modern branch of the calculus of variations, rather well established and mature [39], [68], [46], [76]. Relying on Hamiltonian dynamics is now prevalent, instead of the standard Lagrangian formalism of the calculus of variations. Also, coming from control engineering, constraints on the control (for instance the control is a force or a torque, which are naturally bounded) or the state (for example in the shuttle atmospheric re-entry problem there is a constraint on the thermal flux) are imposed; the ones on the state are usual but these on the state yield more complicated necessary optimality conditions and an increased intrinsic complexity of the optimal solutions. Also, in the modern treatment, ad-hoc numerical schemes have to be derived for effective computations of the optimal solutions.

What makes optimal control an applied field is the necessity of computing these optimal trajectories, or rather the controls that produce these trajectories (or, of course, close-by trajectories). Computing a given optimal trajectory and its control as a function of time is a demanding task, with non trivial numerical difficulties: roughly speaking, the Pontryagin Maximum Principle gives candidate optimal trajectories as solutions of a two point boundary value problem (for an ODE) which can be analyzed using mathematical tools from geometric control theory or solved numerically using shooting methods. Obtaining the *optimal synthesis* –the optimal control as a function of the state– is of course a more intricate problem [46], [51].

These questions are not only academic for minimizing a cost is *very* relevant in many control engineering problems. However, modern engineering textbooks in nonlinear control systems like the “best-seller” [61] hardly mention optimal control, and rather put the emphasis on designing a feedback control, as regular and explicit as possible, satisfying some qualitative (and extremely important!) objectives: disturbance attenuation, decoupling, output regulation or stabilization. Optimal control is sometimes viewed as disconnected from automatic control... we shall come back to this unfortunate point.

3.3.2. Feedback, control Lyapunov functions, stabilization.

A control Lyapunov function (CLF) is a function that can be made a Lyapunov function (roughly speaking, a function that decreases along all trajectories, some call this an “artificial potential”) for the closed-loop system corresponding to *some* feedback law. This can be translated into a partial differential relation sometimes called “Artstein’s (in)equation” [42]. There is a definite parallel between a CLF for stabilization, solution of this differential inequation on the one hand, and the value function of an optimal control problem for the system, solution of a HJB equation on the other hand. Now, optimal control is a quantitative objective while stabilization is a qualitative objective; it is not surprising that Artstein (in)equation is very under-determined and has many more solutions than HJB equation, and that it may (although not always) even have smooth ones.

⁰Consider the simple system with state $(x, y, z) \in \mathbb{R}^3$ and two controls that reads $\dot{z} = (\dot{y} - z\dot{x})^2 \dot{x}$ after elimination of the controls; it is not known whether it is equivalent to a linear system, or flat; this is because the property amounts to existence of a formula giving the general solution as a function of two arbitrary functions of time and their derivatives up to a certain order, but no bound on this order is known a priori, even for this very particular example.

We have, in the team, a longstanding research record on the topic of construction of CLFs and stabilizing feedback controls.

3.4. Optimal Transport

We believe that matching optimal transport with geometric control theory is one originality of our team. We expect interactions in both ways.

The study of optimal mass transport problems in the Euclidean or Riemannian setting has a long history which goes from the pioneer works of Monge [72] and Kantorovitch [65] to the recent revival initiated by fundamental contributions due to Brenier [52] and McCann [71].

The same transportation problems in the presence of differential constraints on the set of paths—like being an admissible trajectory for a control system—is quite new. The first contributors were Ambrosio and Rigot [40] who proved the existence and uniqueness of an optimal transport map for the Monge problem associated with the squared canonical sub-Riemannian distance on the Heisenberg groups. This result was extended later by Agrachev and Lee [37], then by Figalli and Rifford [56] who showed that the Ambrosio-Rigot theorem holds indeed true on many sub-Riemannian manifolds satisfying reasonable assumptions. The problem of existence and uniqueness of an optimal transport map for the squared sub-Riemannian distance on a general complete sub-Riemannian manifold remains open; it is strictly related to the regularity of the sub-Riemannian distance in the product space, and remains a formidable challenge. Generalized notions of Ricci curvatures (bounded from below) in metric spaces have been developed recently by Lott and Villani [69] and Sturm [80]. A pioneer work by Juillet [62] captured the right notion of curvature for subriemannian metric in the Heisenberg group; Agrachev and Lee [38] have elaborated on this work to define new notions of curvatures in three dimensional sub-Riemannian structures. The optimal transport approach happened to be very fruitful in this context. Many things remain to be done in a more general context.

3.5. Small controls and conservative systems, averaging

Using averaging techniques to study small perturbations of integrable Hamiltonian systems dates back to H. Poincaré or earlier; it gives an approximation of the (slow) evolution of quantities that are preserved in the non-perturbed system. It is very subtle in the case of multiple periods but more elementary in the single period case, here it boils down to taking the average of the perturbation along each periodic orbit; see for instance [41], [79].

When the “perturbation” is a control, these techniques may be used after deciding how the control will depend on time and state and other quantities, for instance it may be used after applying the Pontryagin Maximum Principle as in [44], [45], [53], [60]. Without deciding the control a priori, an “average control system” may be defined as in [43].

The focus is then on studying into details this simpler “averaged” problem, that can often be described by a Riemannian metric for quadratic costs or by a Finsler metric for costs like minimum time.

This line of research stemmed out of applications to space engineering, see section 4.1 .

NACHOS Project-Team

3. Research Program

3.1. Scientific foundations

The research activities undertaken by the team aim at developing innovative numerical methodologies putting the emphasis on several features:

- **Accuracy.** The foreseen numerical methods should rely on discretization techniques that best fit to the geometrical characteristics of the problems at hand. Methods based on unstructured, locally refined, even non-conforming, simplicial meshes are particularly attractive in this regard. In addition, the proposed numerical methods should also be capable to accurately describe the underlying physical phenomena that may involve highly variable space and time scales. Both objectives are generally addressed by studying so-called *hp*-adaptive solution strategies which combine *h*-adaptivity using local refinement/coarsening of the mesh and *p*-adaptivity using adaptive local variation of the interpolation order for approximating the solution variables. However, for physical problems involving strongly heterogeneous or high contrast propagation media, such a solution strategy may not be sufficient. Then, for dealing accurately with these situations, one has to design numerical methods that specifically address the multiscale nature of the underlying physical phenomena.
- **Numerical efficiency.** The simulation of unsteady problems most often relies on explicit time integration schemes. Such schemes are constrained by a stability criterion, linking some space and time discretization parameters, that can be very restrictive when the underlying mesh is highly non-uniform (especially for locally refined meshes). For realistic 3D problems, this can represent a severe limitation with regards to the overall computing time. One possible overcoming solution consists in resorting to an implicit time scheme in regions of the computational domain where the underlying mesh size is very small, while an explicit time scheme is applied elsewhere in the computational domain. The resulting hybrid explicit-implicit time integration strategy raises several challenging questions concerning both the mathematical analysis (stability and accuracy, especially for what concern numerical dispersion), and the computer implementation on modern high performance systems (data structures, parallel computing aspects). A second, often considered approach is to devise a local time stepping strategy. Beside, when considering time-harmonic (frequency-domain) wave propagation problems, numerical efficiency is mainly linked to the solution of the system of algebraic equations resulting from the discretization in space of the underlying PDE model. Various strategies exist ranging from the more robust and efficient sparse direct solvers to the more flexible and cheaper (in terms of memory resources) iterative methods. Current trends tend to show that the ideal candidate will be a judicious mix of both approaches by relying on domain decomposition principles.
- **Computational efficiency.** Realistic 3D wave propagation problems involve the processing of very large volumes of data. The latter results from two combined parameters: the size of the mesh i.e the number of mesh elements, and the number of degrees of freedom per mesh element which is itself linked to the degree of interpolation and to the number of physical variables (for systems of partial differential equations). Hence, numerical methods must be adapted to the characteristics of modern parallel computing platforms taking into account their hierarchical nature (e.g multiple processors and multiple core systems with complex cache and memory hierarchies). In addition, appropriate parallelization strategies need to be designed that combine SIMD and MIMD programming paradigms.

From the methodological point of view, the research activities of the team are concerned with four main topics: (1) high order finite element type methods on unstructured or hybrid structured/unstructured meshes for the discretization of the considered systems of PDEs, (2) efficient time integration strategies for dealing with grid induced stiffness when using non-uniform (locally refined) meshes, (3) numerical treatment of complex propagation media models (e.g. physical dispersion models), (4) algorithmic adaptation to modern high performance computing platforms.

3.2. High order discretization methods

3.2.1. The Discontinuous Galerkin method

The Discontinuous Galerkin method (DG) was introduced in 1973 by Reed and Hill to solve the neutron transport equation. From this time to the 90's a review on the DG methods would likely fit into one page. In the meantime, the Finite Volume approach (FV) has been widely adopted by computational fluid dynamics scientists and has now nearly supplanted classical finite difference and finite element methods in solving problems of non-linear convection and conservation law systems. The success of the FV method is due to its ability to capture discontinuous solutions which may occur when solving non-linear equations or more simply, when convecting discontinuous initial data in the linear case. Let us first remark that DG methods share with FV methods this property since a first order FV scheme may be viewed as a 0th order DG scheme. However a DG method may also be considered as a Finite Element (FE) one where the continuity constraint at an element interface is released. While keeping almost all the advantages of the FE method (large spectrum of applications, complex geometries, etc.), the DG method has other nice properties which explain the renewed interest it gains in various domains in scientific computing as witnessed by books or special issues of journals dedicated to this method [41]- [42]- [43]- [48]:

- It is naturally adapted to a high order approximation of the unknown field. Moreover, one may increase the degree of the approximation in the whole mesh as easily as for spectral methods but, with a DG method, this can also be done very locally. In most cases, the approximation relies on a polynomial interpolation method but the DG method also offers the flexibility of applying local approximation strategies that best fit to the intrinsic features of the modeled physical phenomena.
- When the space discretization is coupled to an explicit time integration scheme, the DG method leads to a block diagonal mass matrix whatever the form of the local approximation (e.g. the type of polynomial interpolation). This is a striking difference with classical, continuous FE formulations. Moreover, the mass matrix may be diagonal if the basis functions are orthogonal.
- It easily handles complex meshes. The grid may be a classical conforming FE mesh, a non-conforming one or even a hybrid mesh made of various elements (tetrahedra, prisms, hexahedra, etc.). The DG method has been proven to work well with highly locally refined meshes. This property makes the DG method more suitable (and flexible) to the design of some *hp*-adaptive solution strategy.
- It is also flexible with regards to the choice of the time stepping scheme. One may combine the DG spatial discretization with any global or local explicit time integration scheme, or even implicit, provided the resulting scheme is stable.
- It is naturally adapted to parallel computing. As long as an explicit time integration scheme is used, the DG method is easily parallelized. Moreover, the compact nature of DG discretization schemes is in favor of high computation to communication ratio especially when the interpolation order is increased.

As with standard FE methods, a DG method relies on a variational formulation of the continuous problem at hand. However, due to the discontinuity of the global approximation, this variational formulation has to be defined locally, at the element level. Then, a degree of freedom in the design of a DG method stems from the approximation of the boundary integral term resulting from the application of an integration by parts to the element-wise variational form. In the spirit of FV methods, the approximation of this boundary integral term calls for a numerical flux function which can be based on either a centered scheme or an upwind scheme, or a blending between these two schemes.

3.2.2. High order DG methods for wave propagation models

DG methods are at the heart of the activities of the team regarding the development of high order discretization schemes for the PDE systems modeling electromagnetic and elastodynamic wave propagation.

- **Nodal DG methods for time-domain problems.** For the numerical solution of the time-domain Maxwell equations, we have first proposed a non-dissipative high order DGTD (Discontinuous Galerkin Time-Domain) method working on unstructured conforming simplicial meshes [13]. This DG method combines a central numerical flux function for the approximation of the integral term at the interface of two neighboring elements with a second order leap-frog time integration scheme. Moreover, the local approximation of the electromagnetic field relies on a nodal (Lagrange type) polynomial interpolation method. Recent achievements by the team deal with the extension of these methods towards non-conforming unstructured [10]-[11] and hybrid structured/unstructured meshes [6], their coupling with hybrid explicit/implicit time integration schemes in order to improve their efficiency in the context of locally refined meshes [4]-[19]-[18]. A high order DG method has also been proposed for the numerical resolution of the elastodynamic equations modeling the propagation of seismic waves [2]-[9].
- **Hybridizable DG (HDG) method for time-domain and time-harmonic problems.** For the numerical treatment of the time-harmonic Maxwell equations, nodal DG methods can also be considered [8]. However, such DG formulations are highly expensive, especially for the discretization of 3D problems, because they lead to a large sparse and indefinite linear system of equations coupling all the degrees of freedom of the unknown physical fields. Different attempts have been made in the recent past to improve this situation and one promising strategy has been recently proposed by Cockburn *et al.*[46] in the form of so-called hybridizable DG formulations. The distinctive feature of these methods is that the only globally coupled degrees of freedom are those of an approximation of the solution defined only on the boundaries of the elements. This work is concerned with the study of such Hybridizable Discontinuous Galerkin (HDG) methods for the solution of the system of Maxwell equations in the time-domain when the time integration relies on an implicit scheme, or in the frequency-domain. The team has been a precursor in the development of HDG methods for the frequency-domain Maxwell equations [15]-[16].
- **Multiscale DG methods for time-domain problems.** More recently, in collaboration with LNCC in Petropolis (Frédéric Valentin) the framework of the HOMAR associate team, we are investigating a family of methods specifically designed for an accurate and efficient numerical treatment of multiscale wave propagation problems. These methods, referred to as Multiscale Hybrid Mixed (MHM) methods, are currently studied in the team for both time-domain electromagnetic and elastodynamic PDE models. They consist in reformulating the mixed variational form of each system into a global (arbitrarily coarse) problem related to a weak formulation of the boundary condition (carried by a Lagrange multiplier that represents e.g. the normal stress tensor in elastodynamic systems), and a series of small, element-wise, fully decoupled problems resembling to the initial one and related to some well chosen partition of the solution variables on each element. By construction, that methodology is fully parallelizable and recursivity may be used in each local problem as well, making MHM methods belonging to multi-level highly parallelizable methods. Each local problem may be solved using DG or classical Galerkin FE approximations combined with some appropriate time integration scheme (θ -scheme or leap-frog scheme).

3.3. Efficient time integration strategies

The use of unstructured meshes (based on triangles in two space dimensions and tetrahedra in three space dimensions) is an important feature of the DGTD methods developed in the team which can thus easily deal with complex geometries and heterogeneous propagation media. Moreover, DG discretization methods are naturally adapted to local, conforming as well as non-conforming, refinement of the underlying mesh. Most of the existing DGTD methods rely on explicit time integration schemes and lead to block diagonal mass matrices which is often recognized as one of the main advantages with regards to continuous finite element methods.

However, explicit DGTD methods are also constrained by a stability condition that can be very restrictive on highly refined meshes and when the local approximation relies on high order polynomial interpolation. There are basically three strategies that can be considered to cure this computational efficiency problem. The first approach is to use an unconditionally stable implicit time integration scheme to overcome the restrictive constraint on the time step for locally refined meshes. In a second approach, a local time stepping strategy is combined with an explicit time integration scheme. In the third approach, the time step size restriction is overcome by using a hybrid explicit-implicit procedure. In this case, one blends a time implicit and a time explicit schemes where only the solution variables defined on the smallest elements are treated implicitly. The first and third options are considered in the team in the framework of DG [4]-[19]-[18] and HDG discretization methods.

3.4. Numerical treatment of complex material models

Towards the general aim of being able to consider concrete physical situations, we are interested in taking into account in the numerical methodologies that we study, a better description of the propagation of waves in realistic media. In the case of electromagnetics, a typical physical phenomenon that one has to consider is *dispersion*. It is present in almost all media and expresses the way the material reacts to an electromagnetic field. In the presence of an electric field a medium does not react instantaneously and thus presents an electric polarization of the molecules or electrons that itself influences the electric displacement. In the case of a linear homogeneous isotropic media, there is a linear relation between the applied electric field and the polarization. However, above some range of frequencies (depending on the considered material), the dispersion phenomenon cannot be neglected and the relation between the polarization and the applied electric field becomes complex. This is rendered via a frequency-dependent complex permittivity. Several models of complex permittivity exist. Concerning biological media, the Debye model is commonly adopted in the presence of water, biological tissues and polymers, so that it already covers a wide range of applications [14]. In the context of nanoplasmonics, one is interested in modeling the dispersion effects on metals on the nanometer scale and at optical frequencies. In this case, the Drude or the Drude-Lorentz models are generally chosen [21]. In the context of seismic wave propagation, we are interested by the intrinsic attenuation of the medium [20]. In realistic configurations, for instance in sedimentary basins where the waves are trapped, we can observe site effects due to local geological and geotechnical conditions which result in a strong increase in amplification and duration of the ground motion at some particular locations. During the wave propagation in such media, a part of the seismic energy is dissipated because of anelastic losses related to the internal friction of the medium. For these reasons, numerical simulations based on the basic assumption of linear elasticity are no more valid since this assumption results in a severe overestimation of amplitude and duration of the ground motion, even when we are not in presence of a site effect, since intrinsic attenuation is not taken into account.

3.5. High performance numerical computing

Beside basic research activities related to the design of numerical methods and resolution algorithms for the wave propagation models at hand, the team is also committed to demonstrate the benefits of the proposed numerical methodologies in the simulation of challenging three-dimensional problems pertaining to computational electromagnetics and computational geoseismics. For such applications, parallel computing is a mandatory path. Nowadays, modern parallel computers most often take the form of clusters of heterogeneous multiprocessor systems, combining multiple core CPUs with accelerator cards (e.g Graphical Processing Units - GPUs), with complex hierarchical distributed-shared memory systems. Developing numerical algorithms that efficiently exploit such high performance computing architectures raises several challenges, especially in the context of a massive parallelism. In this context, current efforts of the team are towards the exploitation of multiple levels of parallelism (computing systems combining CPUs and GPUs) through the study of hierarchical SPMD (Single Program Multiple Data) strategies for the parallelization of unstructured mesh based solvers.

TOSCA Project-Team

3. Research Program

3.1. Research Program

Most often physicists, economists, biologists and engineers need a stochastic model because they cannot describe the physical, economical, biological, etc., experiment under consideration with deterministic systems, either because of its complexity and/or its dimension or because precise measurements are impossible. Therefore, they abandon trying to get the exact description of the state of the system at future times given its initial conditions, and try instead to get a statistical description of the evolution of the system. For example, they desire to compute occurrence probabilities for critical events such as the overstepping of a given thresholds by financial losses or neuronal electrical potentials, or to compute the mean value of the time of occurrence of interesting events such as the fragmentation to a very small size of a large proportion of a given population of particles. By nature such problems lead to complex modelling issues: one has to choose appropriate stochastic models, which require a thorough knowledge of their qualitative properties, and then one has to calibrate them, which requires specific statistical methods to face the lack of data or the inaccuracy of these data. In addition, having chosen a family of models and computed the desired statistics, one has to evaluate the sensitivity of the results to the unavoidable model specifications. The TOSCA team, in collaboration with specialists of the relevant fields, develops theoretical studies of stochastic models, calibration procedures, and sensitivity analysis methods.

In view of the complexity of the experiments, and thus of the stochastic models, one cannot expect to use closed form solutions of simple equations in order to compute the desired statistics. Often one even has no other representation than the probabilistic definition (e.g., this is the case when one is interested in the quantiles of the probability law of the possible losses of financial portfolios). Consequently the practitioners need Monte Carlo methods combined with simulations of stochastic models. As the models cannot be simulated exactly, they also need approximation methods which can be efficiently used on computers. The TOSCA team develops mathematical studies and numerical experiments in order to determine the global accuracy and the global efficiency of such algorithms.

The simulation of stochastic processes is not motivated by stochastic models only. The stochastic differential calculus allows one to represent solutions of certain deterministic partial differential equations in terms of probability distributions of functionals of appropriate stochastic processes. For example, elliptic and parabolic linear equations are related to classical stochastic differential equations (SDEs), whereas nonlinear equations such as the Burgers and the Navier–Stokes equations are related to McKean stochastic differential equations describing the asymptotic behavior of stochastic particle systems. In view of such probabilistic representations one can get numerical approximations by using discretization methods of the stochastic differential systems under consideration. These methods may be more efficient than deterministic methods when the space dimension of the PDE is large or when the viscosity is small. The TOSCA team develops new probabilistic representations in order to propose probabilistic numerical methods for equations such as conservation law equations, kinetic equations, and nonlinear Fokker–Planck equations.

ABS Project-Team

3. Research Program

3.1. Introduction

The research conducted by ABS focuses on three main directions in Computational Structural Biology (CSB), together with the associated methodological developments:

- Modeling interfaces and contacts,
- Modeling macro-molecular assemblies,
- Modeling the flexibility of macro-molecules,
- Algorithmic foundations.

3.2. Modeling interfaces and contacts

Keywords: Docking, interfaces, protein complexes, structural alphabets, scoring functions, Voronoi diagrams, arrangements of balls.

The Protein Data Bank, <http://www.rcsb.org/pdb>, contains the structural data which have been resolved experimentally. Most of the entries of the PDB feature isolated proteins⁰, the remaining ones being protein - protein or protein - drug complexes. These structures feature what Nature does – up to the bias imposed by the experimental conditions inherent to structure elucidation, and are of special interest to investigate non-covalent contacts in biological complexes. More precisely, given two proteins defining a complex, interface atoms are defined as the atoms of one protein *interacting* with atoms of the second one. Understanding the structure of interfaces is central to understand biological complexes and thus the function of biological molecules [48]. Yet, in spite of almost three decades of investigations, the basic principles guiding the formation of interfaces and accounting for its stability are unknown [51]. Current investigations follow two routes. From the experimental perspective [34], directed mutagenesis enables one to quantify the energetic importance of residues, important residues being termed *hot* residues. Such studies recently evidenced the *modular* architecture of interfaces [45]. From the modeling perspective, the main issue consists of guessing the hot residues from sequence and/or structural informations [40].

The description of interfaces is also of special interest to improve *scoring functions*. By scoring function, two things are meant: either a function which assigns to a complex a quantity homogeneous to a free energy change⁰, or a function stating that a complex is more stable than another one, in which case the value returned is a score and not an energy. Borrowing to statistical mechanics [29], the usual way to design scoring functions is to mimic the so-called potentials of mean force. To put it briefly, one reverts Boltzmann's law, that is, denoting $p_i(r)$ the probability of two atoms –defining type i – to be located at distance r , the (free) energy assigned to the pair is computed as $E_i(r) = -kT \log p_i(r)$. Estimating from the PDB one function $p_i(r)$ for each type of pair of atoms, the energy of a complex is computed as the sum of the energies of the pairs located within a distance threshold [49], [36]. To compare the energy thus obtained to a reference state, one may compute $E = \sum_i p_i \log p_i/q_i$, with p_i the observed frequencies, and q_i the frequencies stemming from an a priori model [41]. In doing so, the energy defined is nothing but the Kullback-Leibler divergence between the distributions $\{p_i\}$ and $\{q_i\}$.

Describing interfaces poses problems in two settings: static and dynamic.

⁰For structures resolved by crystallography, the PDB contains the asymmetric unit of the crystal. Determining the biological unit from the asymmetric unit is a problem in itself.

⁰The Gibbs free energy of a system is defined by $G = H - TS$, with $H = U + PV$. G is minimum at an equilibrium, and differences in G drive chemical reactions.

In the static setting, one seeks the minimalist geometric model providing a relevant bio-physical signal. A first step in doing so consists of identifying interface atoms, so as to relate the geometry and the bio-chemistry at the interface level [12]. To elaborate at the atomic level, one seeks a structural alphabet encoding the spatial structure of proteins. At the side-chain and backbone level, an example of such alphabet is that of [30]. At the atomic level and in spite of recent observations on the local structure of the neighborhood of a given atom [50], no such alphabet is known. Specific important local conformations are known, though. One of them is the so-called dehydron structure, which is an under-desolvated hydrogen bond – a property that can be directly inferred from the spatial configuration of the C_α carbons surrounding a hydrogen bond [33].

In the dynamic setting, one wishes to understand whether selected (hot) residues exhibit specific dynamic properties, so as to serve as anchors in a binding process [44]. More generally, any significant observation raised in the static setting deserves investigations in the dynamic setting, so as to assess its stability. Such questions are also related to the problem of correlated motions, which we discuss next.

3.3. Modeling macro-molecular assemblies

Keywords: Macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

3.3.1. Reconstruction by Data Integration

Large protein assemblies such as the Nuclear Pore Complex (NPC), chaperonin cavities, the proteasome or ATP synthases, to name a few, are key to numerous biological functions. To improve our understanding of these functions, one would ideally like to build and animate atomic models of these molecular machines. However, this task is especially tough, due to their size and their plasticity, but also due to the flexibility of the proteins involved. In a sense, the modeling challenges arising in this context are different from those faced for binary docking, and also from those encountered for intermediate size complexes which are often amenable to a processing mixing (cryo-EM) image analysis and classical docking. To face these new challenges, an emerging paradigm is that of reconstruction by data integration [28]. In a nutshell, the strategy is reminiscent from NMR and consists of mixing experimental data from a variety of sources, so as to find out the model(s) best complying with the data. This strategy has been in particular used to propose plausible models of the Nuclear Pore Complex [27], the largest assembly known to date in the eukaryotic cell, and consisting of 456 protein *instances* of 30 *types*.

3.3.2. Modeling with Uncertainties and Model Assessment

Reconstruction by data integration requires three ingredients. First, a parametrized model must be adopted, typically a collection of balls to model a protein with pseudo-atoms. Second, as in NMR, a functional measuring the agreement between a model and the data must be chosen. In [26], this functional is based upon *restraints*, namely penalties associated to the experimental data. Third, an optimization scheme must be selected. The design of restraints is notoriously challenging, due to the ambiguous nature and/or the noise level of the data. For example, Tandem Affinity Purification (TAP) gives access to a *pullout* i.e. a list of protein types which are known to interact with one tagged protein type, but no information on the number of complexes or on the stoichiometry of proteins types within a complex is provided. In cryo-EM, the envelope enclosing an assembly is often imprecisely defined, in particular in regions of low density. For immuno-EM labelling experiments, positional uncertainties arise from the microscope resolution.

These uncertainties coupled with the complexity of the functional being optimized, which in general is non convex, have two consequences. First, it is impossible to single out a unique reconstruction, and a set of plausible reconstructions must be considered. As an example, 1000 plausible models of the NPC were reported in [26]. Interestingly, averaging the positions of all balls of a particular protein type across these models resulted in 30 so-called *probability density maps*, each such map encoding the probability of presence of a particular protein type at a particular location in the NPC. Second, the assessment of all models (individual and averaged) is non trivial. In particular, the lack of straightforward statistical analysis of the individual models and the absence of assessment for the averaged models are detrimental to the mechanistic exploitation of the reconstruction results. At this stage, such models therefore remain qualitative.

3.4. Modeling the flexibility of macro-molecules

Keywords: Folding, docking, energy landscapes, induced fit, molecular dynamics, conformers, conformer ensembles, point clouds, reconstruction, shape learning, Morse theory.

Proteins in vivo vibrate at various frequencies: high frequencies correspond to small amplitude deformations of chemical bonds, while low frequencies characterize more global deformations. This flexibility contributes to the entropy thus the free energy of the system *protein - solvent*. From the experimental standpoint, NMR studies generate ensembles of conformations, called conformers, and so do molecular dynamics (MD) simulations. Of particular interest while investigating flexibility is the notion of correlated motion. Intuitively, when a protein is folded, all atomic movements must be correlated, a constraint which gets alleviated when the protein unfolds since the steric constraints get relaxed⁰. Understanding correlations is of special interest to predict the folding pathway that leads a protein towards its native state. A similar discussion holds for the case of partners within a complex, for example in the third step of the *diffusion - conformer selection - induced fit* complex formation model.

Parameterizing these correlated motions, describing the corresponding energy landscapes, as well as handling collections of conformations pose challenging algorithmic problems.

At the side-chain level, the question of improving rotamer libraries is still of interest [32]. This question is essentially a clustering problem in the parameter space describing the side-chains conformations.

At the atomic level, flexibility is essentially investigated resorting to methods based on a classical potential energy (molecular dynamics), and (inverse) kinematics. A molecular dynamics simulation provides a point cloud sampling the conformational landscape of the molecular system investigated, as each step in the simulation corresponds to one point in the parameter space describing the system (the conformational space) [47]. The standard methodology to analyze such a point cloud consists of resorting to normal modes. Recently, though, more elaborate methods resorting to more local analysis [43], to Morse theory [38] and to analysis of meta-stable states of time series [39] have been proposed.

3.5. Algorithmic foundations

Keywords: Computational geometry, computational topology, optimization, data analysis.

Making a stride towards a better understanding of the biophysical questions discussed in the previous sections requires various methodological developments, which we briefly discuss now.

3.5.1. Modeling Interfaces and Contacts

In modeling interfaces and contacts, one may favor geometric or topological information.

On the geometric side, the problem of modeling contacts at the atomic level is tantamount to encoding multi-body relations between an atom and its neighbors. On the one hand, one may use an encoding of neighborhoods based on geometric constructions such as Voronoi diagrams (affine or curved) or arrangements of balls. On the other hand, one may resort to clustering strategies in higher dimensional spaces, as the p neighbors of a given atom are represented by $3p - 6$ degrees of freedom – the neighborhood being invariant upon rigid motions. The information gathered while modeling contacts can further be integrated into interface models.

On the topological side, one may favor constructions which remain stable if each atom in a structure *retains* the same neighbors, even though the 3D positions of these neighbors change to some extent. This process is observed in flexible docking cases, and call for the development of methods to encode and compare shapes undergoing tame geometric deformations.

3.5.2. Modeling Macro-molecular Assemblies

In dealing with large assemblies, a number of methodological developments are called for.

⁰Assuming local forces are prominent, which in turn subsumes electrostatic interactions are not prominent.

On the experimental side, of particular interest is the disambiguation of proteomics signals. For example, TAP and mass spectrometry data call for the development of combinatorial algorithms aiming at unraveling pairwise contacts between proteins within an assembly. Likewise, density maps coming from electron microscopy, which are often of intermediate resolution (5-10Å) call the development of noise resilient segmentation and interpretation algorithms. The results produced by such algorithms can further be used to guide the docking of high resolutions crystal structures into maps.

As for modeling, two classes of developments are particularly stimulating. The first one is concerned with the design of algorithms performing reconstruction by data integration, a process reminiscent from non convex optimization. The second one encompasses assessment methods, in order to single out the reconstructions which best comply with the experimental data. For that endeavor, the development of geometric and topological models accommodating uncertainties is particularly important.

3.5.3. Modeling the Flexibility of Macro-molecules

Given a sampling on an energy landscape, a number of fundamental issues actually arise: how does the point cloud describe the topography of the energy landscape (a question reminiscent from Morse theory)? Can one infer the effective number of degrees of freedom of the system over the simulation, and is this number varying? Answers to these questions would be of major interest to refine our understanding of folding and docking, with applications to the prediction of structural properties. It should be noted in passing that such questions are probably related to modeling phase transitions in statistical physics where geometric and topological methods are being used [42].

From an algorithmic standpoint, such questions are reminiscent of *shape learning*. Given a collection of samples on an (unknown) *model*, *learning* consists of guessing the model from the samples – the result of this process may be called the *reconstruction*. In doing so, two types of guarantees are sought: topologically speaking, the reconstruction and the model should (ideally!) be isotopic; geometrically speaking, their Hausdorff distance should be small. Motivated by applications in Computer Aided Geometric Design, surface reconstruction triggered a major activity in the Computational Geometry community over the past ten years [7]. Aside from applications, reconstruction raises a number of deep issues: the study of distance functions to the model and to the samples, and their comparison; the study of Morse-like constructions stemming from distance functions to points; the analysis of topological invariants of the model and the samples, and their comparison.

ASCLEPIOS Project-Team

3. Research Program

3.1. Introduction

Tremendous progress has been made in the automated analysis of biomedical images during the past two decades [72]. Readers who are neophytes to the field of medical imaging will find an interesting presentation of acquisition techniques of the main medical imaging modalities in [64], [62]. Regarding target applications, a good review of the state of the art can be found in the book *Computer Integrated Surgery* [60], in N. Ayache's article [67] and in recent review articles [68], [72]. The scientific journals *Medical Image Analysis* [55], *Transactions on Medical Imaging* [61], and *Computer Assisted Surgery* [63] are also good reference material. One can have a good vision of the state of the art from the proceedings of the MICCAI'2010 (Medical Image Computing and Computer Assisted Intervention [58], [59]) and ISBI'2010 (Int. Symp. on Biomedical Imaging [57]) conferences.

For instance, for rigid parts of the body like the head, it is now possible to fuse in a completely automated manner images of the same patient taken from different imaging modalities (e.g. anatomical and functional), or to track the evolution of a pathology through the automated registration and comparison of a series of images taken at distant time instants [73], [83]. It is also possible to obtain from a Magnetic Resonance Image (MRI) of the head a reasonable segmentation of skull tissues, white matter, grey matter, and cerebro-spinal fluid [86], or to measure some functional properties of the heart from dynamic sequences of Magnetic Resonance [66], Ultrasound or Nuclear Medicine images [74].

Despite these advances and successes, statistical models of anatomy are still very crude, resulting in poor registration results in deformable regions of the body, or between different subjects. If some algorithms exploit the physical modeling of the image acquisition process, only a few actually model the physical or even the physiological properties of the human body itself. Coupling biomedical image analysis with anatomical and physiological models of the human body could not only provide a better understanding of observed images and signals, but also more efficient tools for detecting anomalies, predicting evolutions, simulating and assessing therapies.

3.2. Medical Image Analysis

The quality of biomedical images tends to improve constantly (better spatial and temporal resolution, better signal to noise ratio). Not only are the images multidimensional (3 spatial coordinates and possibly one temporal dimension), but medical protocols tend to include multisequence (or multiparametric)⁰ and multimodal images⁰ for each single patient.

⁰Multisequence (or multiparametric) imaging consists in acquiring several images of a given patient with the same imaging modality (e.g. MRI, CT, US, SPECT, etc.) but with varying acquisition parameters. For instance, using MRI, patients followed for multiple sclerosis may undergo every six months a 3D multisequence MR acquisition protocol with different pulse sequences (called T1, T2, PD, Flair, etc.): by varying some parameters of the pulse sequences (e.g. Echo Time and Repetition Time), images of the same regions are produced with quite different contrasts depending on the nature and function of the observed structures. In addition, one of the acquisitions (T1) can be combined with the injection of a contrast product (typically Gadolinium) to reveal vessels and some pathologies. Diffusion Tensor Images (DTI) can be acquired to measure the self diffusion of protons in every voxel, allowing the measurement for instance of the direction of white matter fibers in the brain (the same principle can be used to measure the direction of muscular fibers in the heart). Functional MRI of the brain can be acquired by exploiting the so-called Bold Effect (Blood Oxygen Level Dependency): slightly higher blood flow in active regions creates a subtle higher T2* signal which can be detected with sophisticated image processing techniques.

⁰Multimodal acquisition consists in acquiring from the same patient images of different modalities, in order to exploit their complementary nature. For instance, CT and MR may provide information on the anatomy (CT providing contrast between bones and soft tissues while MR within soft tissues of different nature) while SPECT and PET images may provide functional information by measuring a local level of metabolic activity.

Despite remarkable efforts and advances during the past twenty years, the central problems of segmentation and registration have not been solved in the general case. It is our objective in the short term to work on specific versions of these problems, taking into account as much *a priori* information as possible on the underlying anatomy and pathology at hand. It is also our objective to include more knowledge of the physics of image acquisition and observed tissues, as well as of the biological processes involved. Therefore the research activities mentioned in this section will incorporate the advances made in Computational Anatomy and Computational Physiology, as described in sections 3.3 and 3.4 .

We plan to pursue our efforts on the following problems:

- multi-dimensional, multi-sequence and multi-modal image segmentation; and
- image Registration/Fusion.

3.3. Computational Anatomy

The aim of Computational Anatomy (CA) is to model and analyse the biological variability of the human anatomy. Typical applications cover the simulation of average anatomies and normal variations, the discovery of structural differences between healthy and diseased populations, and the detection and classification of pathologies from structural anomalies.⁰

Studying the variability of biological shapes is an old problem (cf. the book "On Shape and Growth" by D'Arcy Thompson [85]). Significant efforts have since been made to develop a theory for statistical shape analysis (one can refer to [71] for a good summary, and to the special issue of Neuroimage [84] for recent developments). Despite all these efforts, there are a number of challenging mathematical issues that remain largely unsolved. A particular issue is the computation of statistics on manifolds that can be of infinite dimension (e.g the group of diffeomorphisms).

There is a classical stratification of the problems into the following 3 levels [80]:

1. construction from medical images of anatomical manifolds of points, curves, surfaces and volumes;
2. assignment of a point to point correspondence between these manifolds using a specified class of transformations (e.g. rigid, affine, diffeomorphism);
3. generation of probability laws of anatomical variation from these correspondences.

We plan to focus our efforts on the following problems:

1. statistics on anatomical manifolds;
2. propagation of variability from anatomical manifolds;
3. linking anatomical variability to image analysis algorithms; and
4. grid-computing strategies to exploit large databases.

3.4. Computational Physiology

The objective of Computational Physiology (CP) is to provide models of the major functions of the human body and numerical methods to simulate them. The main applications are in medicine where CP can for instance be used to better understand the basic processes leading to the appearance of a pathology, to model its probable evolution and to plan, simulate, and monitor its therapy.

⁰The NIH has launched in 2005 the Alzheimer's Disease Neuroimaging Initiative (60 million USD), a multi-center MRI study of 800 patients who will be followed during several years. The aim is to establish new surrogate end-points from the automated analysis of temporal sequences, which is a challenging goal for researchers in Computational Anatomy. The data is to be made available to qualified research groups involved or not in the study.

Quite advanced models have already been proposed to study at the molecular, cellular and organ level a number of physiological systems (see for instance [81], [78], [69], [82], [75]). While these models and new ones need to be developed, refined or validated, a grand challenge that we want to address in this project is the automatic adaptation of the model to a given patient by comparing the model with the available biomedical images and signals and possibly also some additional information (e.g. genetic). Building such *patient-specific models* is an ambitious goal, which requires the choice or construction of models with a complexity adapted to the resolution of the accessible measurements and the development of new data assimilation methods coping with massive numbers of measurements and unknowns.

There is a hierarchy of modeling levels for CP models of the human body [70]:

- the first level is mainly geometrical, and addresses the construction of a digital description of the anatomy [65], essentially acquired from medical imagery;
- the second level is physical, involving mainly the biomechanical modeling of various tissues, organs, vessels, muscles and bone structures [76];
- the third level is physiological, involving the modeling of the functions of the major organ systems [77] (e.g. cardiovascular, respiratory, digestive, central or peripheral nervous, muscular, reproductive, hormonal) or some pathological metabolism (e.g. evolution of cancerous or inflammatory lesions, formation of vessel stenoses, etc.); and
- a fourth level is cognitive, modeling the higher functions of the human brain [56].

These different levels of modeling are closely related to each other, and several physiological systems may interact with each other (e.g. the cardiopulmonary interaction [79]). The choice of the resolution at which each level is described is important, and may vary from microscopic to macroscopic, ideally through multiscale descriptions.

Building this complete hierarchy of models is necessary to evolve from a *Visible Human project* (essentially the first level of modeling) to a much more ambitious *Physiological Human project* (see [77], [78]). We will not address all the issues raised by this ambitious project, but instead focus on the topics detailed below. Among them, our objective is to identify some common methods for the resolution of the large inverse problem raised by the coupling of physiological models and medical images for the construction of patient-specific models (e.g. specific variational or sequential methods (EKF), dedicated particle filters). We also plan to develop specific expertise in the extraction of geometrical meshes from medical images for their further use in simulation procedures. Finally, computational models can be used for specific image analysis problems studied in section 3.2 (e.g. segmentation, registration, tracking). Application domains include

1. surgery simulation;
2. cardiac Imaging;
3. brain tumors, neo-angiogenesis, wound healing processes, ovocyte regulation, etc.

3.5. Clinical Validation

If the objective of many of the research activities of the project is the discovery of original methods and algorithms with a proof of its feasibility in a limited number of representative cases (i.e. proofs of concept) and publications in high quality scientific journals, we believe that it is important that a reasonable number of studies include a much more significant validation effort. As the BioMedical Image Analysis discipline becomes more mature, validation is necessary for the transformation of new ideas into clinical tools and/or industrial products. It also helps to get access to larger databases of images and signals, which in turn help to stimulate new ideas and concepts.

ATHENA Project-Team

3. Research Program

3.1. Computational diffusion MRI

Diffusion MRI (dMRI) provides a non-invasive way of estimating in-vivo CNS fiber structures using the average random thermal movement (diffusion) of water molecules as a probe. It's a recent field of research with a history of roughly three decades. It was introduced in the mid 80's by Le Bihan et al [90], Merboldt et al [94] and Taylor et al [103]. As of today, it is the unique non-invasive technique capable of describing the neural connectivity in vivo by quantifying the anisotropic diffusion of water molecules in biological tissues.

3.1.1. Diffusion Tensor Imaging & High Angular Resolution Diffusion Imaging

In dMRI, the acquisition and reconstruction of the diffusion signal allows for the reconstruction of the water molecules displacement probability, known as the Ensemble Average Propagator (EAP) [102], [72]. Historically, the first model in dMRI is the 2nd order diffusion tensor (DTI) [70], [69] which assumes the EAP to be Gaussian centered at the origin. DTI has now proved to be extremely useful to study the normal and pathological human brain [91], [80]. It has led to many applications in clinical diagnosis of neurological diseases and disorder, neurosciences applications in assessing connectivity of different brain regions, and more recently, therapeutic applications, primarily in neurosurgical planning. An important and very successful application of diffusion MRI has been brain ischemia, following the discovery that water diffusion drops immediately after the onset of an ischemic event, when brain cells undergo swelling through cytotoxic edema.

The increasing clinical importance of diffusion imaging has driven our interest to develop new processing tools for Diffusion Tensor MRI. Because of the complexity of the data, this imaging modality raises a large amount of mathematical and computational challenges. We have therefore developed original and efficient algorithms relying on Riemannian geometry, differential geometry, partial differential equations and front propagation techniques to correctly and efficiently estimate, regularize, segment and process Diffusion Tensor MRI (DT-MRI) (see [93] and [92]).

In DTI, the Gaussian assumption over-simplifies the diffusion of water molecules. While it is adequate for voxels in which there is only a single fiber orientation (or none), it breaks for voxels in which there are more complex internal structures and limitates the ability of the DTI to describe complex, singular and intricate fiber configurations (U-shape, kissing or crossing fibers). To overcome this limitation, so-called Diffusion Spectrum Imaging (DSI) [107] and High Angular Resolution Diffusion Imaging (HARDI) methods such as Q-ball imaging [105] and other multi-tensors and compartment models [100], [101], [63], [62], [98] were developed to resolve the orientationality of more complicated fiber bundle configurations.

Q-Ball imaging (QBI) has been proven very successful in resolving multiple intravoxel fiber orientations in MR images, thanks to its ability to reconstruct the Orientation Distribution Function (ODF, the probability of diffusion in a given direction). These tools play a central role in our work related to the development of a robust and linear spherical harmonic estimation of the HARDI signal and to our development of a regularized, fast and robust analytical QBI solution that outperforms the state-of-the-art ODF numerical technique developed by Tuch. Those contributions are fundamental and have already started to impact on the Diffusion MRI, HARDI and Q-Ball Imaging community [79]. They are at the core of our probabilistic and deterministic tractography algorithms devised to best exploit the full distribution of the fiber ODF (see [76], [5] and [77],[6]).

3.1.2. Beyond DTI with high order tensors

High Order Tensors (HOT) models to estimate the diffusion function while overcoming the shortcomings of the 2nd order tensor model have also been recently proposed such as the Generalized Diffusion Tensor Imaging (G-DTI) model developed by Ozarslan et al [109], [110] or 4th order Tensor Model [68]. For more details, we refer the reader to our articles in [81], [100] where we review HOT models and to our articles

in [92], co-authored with some of our close collaborators, where we review recent mathematical models and computational methods for the processing of Diffusion Magnetic Resonance Images, including state-of-the-art reconstruction of diffusion models, cerebral white matter connectivity analysis, and segmentation techniques. Recently, we started to work on Diffusion Kurtosis Imaging (DKI), of great interest for the company OLEA MEDICAL. Indeed, DKI is fast gaining popularity in the domain for characterizing the diffusion propagator or EAP by its deviation from Gaussianity. Hence it is an important tool in the clinic for characterizing the white-matter's integrity with biomarkers derived from the 3D 4th order kurtosis tensor (KT) [84].

All these powerful techniques are of utmost importance to acquire a better understanding of the CNS mechanisms and have helped to efficiently tackle and solve a number of important and challenging problems [62], [63]. They have also opened up a landscape of extremely exciting research fields for medicine and neuroscience. Hence, due to the complexity of the CNS data and as the magnetic field strength of scanners increase, as the strength and speed of gradients increase and as new acquisition techniques appear [4], these imaging modalities raise a large amount of mathematical and computational challenges at the core of the research we develop at ATHENA [83], [100].

3.1.3. Improving dMRI acquisitions

One of the most important challenges in diffusion imaging is to improve acquisition schemes and analyse approaches to optimally acquire and accurately represent diffusion profiles in a clinically feasible scanning time. Indeed, a very important and open problem in Diffusion MRI is related to the fact that HARDI scans generally require many times more diffusion gradient than traditional diffusion MRI scan times. This comes at the price of longer scans, which can be problematic for children and people with certain diseases. Patients are usually unable to tolerate long scans and excessive motion of the patient during the acquisition process can force a scan to be aborted or produce useless diffusion MRI images. Recently, we have developed novel methods for the acquisition and the processing of diffusion magnetic resonance images, to efficiently provide, with just few measurements, new insights into the structure and anatomy of the brain white matter in vivo.

First, we contributed developing real-time reconstruction algorithm based on the Kalman filter [75]. Then, and more recently, we started to explore the utility of Compressive Sensing methods to enable faster acquisition of dMRI data by reducing the number of measurements, while maintaining a high quality for the results. Compressed Sensing (CS) is a recent technique which has been proved to accurately reconstruct sparse signals from undersampled measurements acquired below the Shannon-Nyquist rate [95].

We have contributed to the reconstruction of the diffusion signal and its important features as the orientation distribution function and the ensemble average propagator, with a special focus on clinical setting in particular for single and multiple Q-shell experiments [95], [73], [74]. Compressive sensing as well as the parametric reconstruction of the diffusion signal in a continuous basis of functions such as the Spherical Polar Fourier basis, have been proved through our recent contributions to be very useful for deriving simple and analytical closed formulae for many important dMRI features, which can be estimated via a reduced number of measurements [95], [73], [74].

We have also contributed to design optimal acquisition schemes for single and multiple q-shell experiments. In particular, the method proposed in [4] helps generate sampling schemes with optimal angular coverage for multi-shell acquisitions. The cost function we proposed is an extension of the electrostatic repulsion to multi-shell and can be used to create acquisition schemes with incremental angular distribution, compatible with prematurely stopped scans. Compared to more commonly used radial sampling, our method improves the angular resolution, as well as fiber crossing discrimination. The optimal sampling schemes, freely available for download⁰, have been selected for use in the HCP (Human Connectome Project)⁰.

We think that such kind of contributions open new perspectives for dMRI applications including, for example, tractography where the improved characterization of the fiber orientations is likely to greatly and quickly help tracking through regions with and/or without crossing fibers [82]

⁰<http://www.emmanuelcaruyer.com/>

⁰<http://humanconnectome.org/documentation/Q1/imaging-protocols.html>

3.1.4. *dMRI modelling, tissue microstructures features recovery & applications*

The dMRI signal is highly complex, hence, the mathematical tools required for processing it have to be commensurate in their complexity. Overall, these last twenty years have seen an explosion of intensive scientific research which has vastly improved and literally changed the face of dMRI. In terms of dMRI models, two trends are clearly visible today: the parametric approaches which attempt to build models of the tissue to explain the signal based on model-parameters such as CHARMED [64], AxCaliber [65] and NODDI [108] to cite but a few, and the non-parametric approaches, which attempt to describe the signal in useful but generic functional bases such as the Spherical Polar Fourier (SPF) basis [67], [66], the Solid Harmonic (SoH) basis [78], the Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE) basis [96] and more recent Mean Apparent Propagator or MAP-MRI basis [97].

However, although great improvements have been made in the last twenty years, major improvements are still required primarily to optimally acquire dMRI data, better understand the biophysics of the signal formation, recover invariant and intrinsic microstructure features, identify bio-physically important bio-markers and improve tractography. For short, there is still considerable room for improvement to take dMRI from the benchside to the bedside.

Therefore, there is still considerable room for improvement when it comes to the concepts and tools able to efficiently acquire, process and analyze the complex structure of dMRI data. Develop ground-breaking tools and models for dMRI is one of the major objectives we would like to achieve in order to lead to a decisive advance and breakthrough in this field.

Then, we propose to investigate the feasibility of using our new models and methods to measure extremely important biological tissue microstructure quantities such as axonal radius and density in white matter. These parameters could indeed provide new insight to better understand the brain's architecture and more importantly could also provide new imaging bio-markers to characterize certain neurodegenerative diseases. This challenging scientific problem, when solved, will lead to direct measurements of important microstructural features that will be integrated in our analysis to provide much greater insight into disease mechanisms, recovery and development. These new microstructural parameters will open the road to go far beyond the limitations of the more simple bio-markers derived from DTI that are clinically used to this date – such as MD and FA which are known to be extremely sensitive to confounding factors such as partial volume and axonal dispersion, non-specific and not able to capture any subtle effects that might be early indicators of diseases [7].

3.1.5. *Towards microstructural based tractography*

In order to go far beyond traditional fiber-tracking techniques, we believe that first order information, i.e. fiber orientations, has to be superseded by second and third order information, such as microstructure details, to improve tractography. However, many of these higher order information methods are relatively new or unexplored and tractography algorithms based on these high order based methods have to be conceived and designed. In this aim, we propose to work with multiple-shells to reconstruct the Ensemble Average Propagator (EAP), which represents the whole 3D diffusion process and use the possibility it offers to deduce valuable insights on the microstructural properties of the white matter. Indeed, from a reconstructed EAP one can compute the angular features of the diffusion in an diffusion Orientation Distribution Function (ODF), providing insight in axon orientation, calculate properties of the entire diffusion in a voxel such as the Mean Squared Diffusivity (MSD) and Return-To-Origin Probability (RTOP), or come forth with bio-markers detailing diffusion along a particular white matter bundle direction such as the Return-to-Axis or Return-to-Plane Probability (RTAP or RTPP). This opens the way to a ground-breaking computational and unified framework for tractography based on EAP and microstructure features [8]. Using additional a priori anatomical [11] and/or functional information, we could also constrain the tractography algorithm to start and terminate the streamlines only at valid processing areas of the brain.

This development of a computational and unified framework for tractography, based on EAP, microstructure and a priori anatomical and/or functional features, will open new perspectives in tractography, paving the way to a new generation of realistic and biologically plausible algorithms able to deal with intricate configurations of white matter fibers and to provide an exquisite and intrinsic brain connectivity quantification.

3.2. MEG and EEG

Electroencephalography (EEG) and Magnetoencephalography (MEG) are two non-invasive techniques for measuring (part of) the electrical activity of the brain. While EEG is an old technique (Hans Berger, a German neuropsychiatrist, measured the first human EEG in 1929), MEG is a rather new one: the first measurements of the magnetic field generated by the electrophysiological activity of the brain were made in 1968 at MIT by D. Cohen. Nowadays, EEG is relatively inexpensive and is routinely used to detect and qualify neural activities (epilepsy detection and characterisation, neural disorder qualification, BCI, ...). MEG is, comparatively, much more expensive as SQUIDS only operate under very challenging conditions (at liquid helium temperature) and as a specially shielded room must be used to separate the signal of interest from the ambient noise. However, as it reveals a complementary vision to that of EEG and as it is less sensitive to the head structure, it also bears great hopes and an increasing number of MEG machines are being installed throughout the world. Inria and ODYSSEÉ/ATHENA have participated in the acquisition of one such machine installed in the hospital "La Timone" in Marseille.

MEG and EEG can be measured simultaneously (M/EEG) and reveal complementary properties of the electrical fields. The two techniques have temporal resolutions of about the millisecond, which is the typical granularity of the measurable electrical phenomena that arise within the brain. This high temporal resolution makes MEG and EEG attractive for the functional study of the brain. The spatial resolution, on the contrary, is somewhat poor as only a few hundred data points can be acquired simultaneously (about 300-400 for MEG and up to 256 for EEG). MEG and EEG are somewhat complementary with fMRI and SPECT in that those provide a very good spatial resolution but a rather poor temporal resolution (of the order of a second for fMRI and a minute for SPECT). Also, contrarily to fMRI, which "only" measures an haemodynamic response linked to the metabolic demand, MEG and EEG measure a direct consequence of the electrical activity of the brain: it is acknowledged that the signals measured by MEG and EEG correspond to the variations of the post-synaptic potentials of the pyramidal cells in the cortex. Pyramidal neurons compose approximately 80% of the neurons of the cortex, and it requires at least about 50,000 active such neurons to generate some measurable signal.

While the few hundred temporal curves obtained using M/EEG have a clear clinical interest, they only provide partial information on the localisation of the sources of the activity (as the measurements are made on or outside of the head). Thus the practical use of M/EEG data raises various problems that are at the core of the ATHENA research in this topic:

- First, as acquisition is continuous and is run at a rate up to 1kHz, the amount of data generated by each experiment is huge. Data selection and reduction (finding relevant time blocks or frequency bands) and pre-processing (removing artifacts, enhancing the signal to noise ratio, ...) are largely done manually at present. Making a better and more systematic use of the measurements is an important step to optimally exploit the M/EEG data [3].
- With a proper model of the head and of the sources of brain electromagnetic activity, it is possible to simulate the electrical propagation and reconstruct sources that can explain the measured signal. Proposing better models [89], [10] and means to calibrate them [106] so as to have better reconstructions are other important aims of our work.
- Finally, we wish to exploit the temporal resolution of M/EEG and to apply the various methods we have developed to better understand some aspects of the brain functioning, and/or to extract more subtle information out of the measurements. This is of interest not only as a cognitive goal, but it also serves the purpose of validating our algorithms and can lead to the use of such methods in the field of Brain Computer Interfaces. To be able to conduct such kind of experiments, an EEG lab has been set up at ATHENA.

BIOCORE Project-Team

3. Research Program

3.1. Mathematical and computational methods

BIOCORE's action is centered on the mathematical modeling of biological systems, more particularly of artificial ecosystems, that have been built or strongly shaped by human. Indeed, the complexity of such systems where life plays a central role often makes them impossible to understand, control, or optimize without such a formalization. Our theoretical framework of choice for that purpose is Control Theory, whose central concept is "the system", described by state variables, with inputs (action on the system), and outputs (the available measurements on the system). In modeling the ecosystems that we consider, mainly through ordinary differential equations, the state variables are often population, substrate and/or food densities, whose evolution is influenced by the voluntary or involuntary actions of man (inputs and disturbances). The outputs will be some product that one can collect from this ecosystem (harvest, capture, production of a biochemical product, etc), or some measurements (number of individuals, concentrations, etc). Developing a model in biology is however not straightforward: the absence of rigorous laws as in physics, the presence of numerous populations and inputs in the ecosystems, most of them being irrelevant to the problem at hand, the uncertainties and noise in experiments or even in the biological interactions require the development of dedicated techniques to identify and validate the structure of models from data obtained by or with experimentalists.

Building a model is rarely an objective in itself. Once we have checked that it satisfies some biological constraints (eg. densities stay positive) and fitted its parameters to data (requiring tailor-made methods), we perform a mathematical analysis to check that its behavior is consistent with observations. Again, specific methods for this analysis need to be developed that take advantage of the structure of the model (eg. the interactions are monotone) and that take into account the strong uncertainty that is linked to life, so that qualitative, rather than quantitative, analysis is often the way to go.

In order to act on the system, which often is the purpose of our modeling approach, we then make use of two strong points of Control Theory: 1) the development of observers, that estimate the full internal state of the system from the measurements that we have, and 2) the design of a control law, that imposes to the system the behavior that we want to achieve, such as the regulation at a set point or optimization of its functioning. However, due to the peculiar structure and large uncertainties of our models, we need to develop specific methods. Since actual sensors can be quite costly or simply do not exist, a large part of the internal state often needs to be re-constructed from the measurements and one of the methods we developed consists in integrating the large uncertainties by assuming that some parameters or inputs belong to given intervals. We then developed robust observers that asymptotically estimate intervals for the state variables [73]. Using the directly measured variables and those that have been obtained through such, or other, observers, we then develop control methods that take advantage of the system structure (linked to competition or predation relationships between species in bioreactors or in the trophic networks created or modified by biological control).

3.2. A methodological approach to biology: from genes to ecosystems

One of the objectives of BIOCORE is to develop a methodology that leads to the integration of the different biological levels in our modeling approach: from the biochemical reactions to ecosystems. The regulatory pathways at the cellular level are at the basis of the behavior of the individual organism but, conversely, the external stresses perceived by the individual or population will also influence the intracellular pathways. In a modern "systems biology" view, the dynamics of the whole biosystem/ecosystem emerge from the interconnections among its components, cellular pathways/individual organisms/population. The different scales of size and time that exist at each level will also play an important role in the behavior of the biosystem/ecosystem. We intend to develop methods to understand the mechanisms at play at each level,

from cellular pathways to individual organisms and populations; we assess and model the interconnections and influence between two scale levels (eg., metabolic and genetic; individual organism and population); we explore the possible regulatory and control pathways between two levels; we aim at reducing the size of these large models, in order to isolate subsystems of the main players involved in specific dynamical behaviors.

We develop a theoretical approach of biology by simultaneously considering different levels of description and by linking them, either bottom up (scale transfer) or top down (model reduction). These approaches are used on modeling and analysis of the dynamics of populations of organisms; modeling and analysis of small artificial biological systems using methods of systems biology; control and design of artificial and synthetic biological systems, especially through the coupling of systems.

The goal of this multi-level approach is to be able to design or control the cell or individuals in order to optimize some production or behavior at higher level: for example, control the growth of microalgae via their genetic or metabolic networks, in order to optimize the production of lipids for bioenergy at the photobioreactor level.

BIOVISION Team

3. Research Program

3.1. Introduction

The Biovision team has started on January 1st, 2016. It aims at developing fundamental research as well as technological developments along two axes.

3.1.1. Axis 1: High tech vision aid systems for low vision patients

The most popular class of vision aid systems for low vision patients is based on the idea of magnification. These aids are helpful for tasks such as reading but of course are not useful in other common daily tasks such as navigation.

Video goggles⁰ are another kind of device where visual information is captured by a head-mounted camera, processed and then displayed on a near-the-eye display screen. So far, this technology did not encounter a big success essentially due to their narrow field of view. This situation could evolve with the fast progression of technology around virtual reality and augmented reality.

In BIOVISION we mainly focus on this technology to develop new vision aid systems that could take into account the pathologies of low vision patients but also on the tasks performed by the patients. We have three main goals:

1. We plan to focus on three tasks: reading, watching movies and navigating (indoor or outdoor), which are all important daily life activities for patients.
2. We aim at proposing new **scene enhancements** depending on **pathologies**.
3. We want to test them in **immersive** environments with low vision patients, taking into consideration **ergonomics**.

3.1.2. Axis 2: Human vision understanding through joint experimental and modeling studies, for normal and dystrophic retinas

A holistic point of view is emerging in neuroscience where one can observe simultaneously how vision works at different levels of the hierarchy in the visual system. Multiple scales functional analysis and connectomics are also exploding in brain science, and studies of visual systems are upfront on this fast move. These integrated studies call for new classes of theoretical and integrated models where the goal is the modeling of visual functions such as motion integration.

In BIOVISION we contribute to a better understanding of the visual system with three main goals:

1. We aim at proposing simplified mathematical models characterizing how the **retina** converts a visual scene into spike **population coding**, in **normal and under specific pathological conditions**.
2. We want to design an integrated numerical model of the visual stream, with a focus on motion integration, from retina to **visual cortex** area (e.g., the motion stream **V1-MT-MST**).
3. We plan to develop a simulation platform emulating the retinal spike-response to visual and prosthetic simulations, in normal and pathological conditions.

Finally, although this is not the main goal of our team, another natural avenue of our research will be to develop novel synergistic solutions to solve computer vision tasks based on bio-inspired mechanisms.

3.2. Scientific methodology

In this section we briefly describe the scientific methods we use to achieve our research goals.

⁰Video goggles are marketed by several companies such as, e.g., eSight, Enhanced Vision and Lumus

3.2.1. Adaptive image processing

An impressive range of techniques have been developed in the fields of image processing, computer vision and computer graphics to manipulate and interpret image content for a variety of applications. So far only a few of these techniques have been applied in the context of vision aid systems and even less have been carefully evaluated with patients. However it is worth noticing a recent gain of interest from the artificial vision side to low vision applications⁰. We investigate which techniques could bring a real interest for vision aid systems, how to combine them and how to make them adapted to patient needs, so that they can not only "see" an image but understand it more efficiently.

Some techniques have already been explored. Among the first, enhancing image content (equalization, gamma correction, tone mapping, edge enhancement, image decomposition, cartoonization) seems a natural type of processing to make. Some methods have already been tested with low vision patients [38], [54], [55] or even in retina prosthesis systems as a pre-processing [37]. For some visual impairment it can be useful to consider methods that help patients to focus on the most relevant information, using techniques such as scene retargeting [59], seam carving [40], [39], saliency-based enhancements [71], [82] or 3D-based enhancements when available [64]. All the work done on image understanding could also be extremely useful to help patients navigate in natural cluttered environments both in low vision condition or for prosthetics vision [58]. 3D information, obtained from stereo head systems or RGB-D cameras also bring useful information about the environment [62] and integrated systems combining different expertise are appearing [46].

Our goal will be to take the most of state-of-the-art computer vision methods, in combination with virtual and augmented reality devices (Sec. 3.2.2) to provide patients vision aid system that can adapt to their impairment and so that they can easily change the parameters of the processing in an intuitive way.

3.2.2. Virtual and augmented reality

Our goal is to develop vision-aid systems using virtual and augmented reality [87]. There is a rich continuum of devices between virtual reality (which is *a priori* simpler to use since there is no problem of mobility and environment is well defined), and augmented reality (where information has to be superimposed in real time on top of the real environment to enrich it). Between these two extremes, new hybrid see-through systems are available or under development such as light glasses where additional information can be locally displayed at the center or on the corner (e.g., Google glass improving it). We invest on these technologies which enable new kinds of interaction with visual content which could be very powerful when adapted to low vision patients who want to use their remaining sight. We investigate how low vision patients could take benefits from this technology in their daily life activities [47]⁰.

We focus on three activities: reading, watching movies and navigating in real world (indoor and outdoor). In these three scenario, this technology should offer crucial advantages for people in low vision. For reading, this could help them solving the page navigation problem or the limitations of magnification encountered when standard CCTVs are used. When watching a movie, the possibility to explore a pre-processed visual scene presented with very high visual angle can help patients to follow the storyline more easily and this poses some interesting questions on the creation of content specifically for virtual reality headsets. Finally, in real scenarios, augmented reality offers promising perspectives to enrich the scene by highly visible visual cues to facilitate low vision patients navigation. Of course the choices of adaptive image processing techniques (see Sec. 3.2.1) will be crucial and this will be the add-on value of our work.

⁰See, e.g., the [Special issue on Assistive Computer Vision and Robotics - "Assistive Solutions for Mobility, Communication and HMI"](#) from Computer Vision and Image Understanding (August 2016) or the International Workshop on Assistive Computer Vision and Robotics ([ECCV 2016 Sattelite workshop](#))

⁰Note that wearing such headsets may not be easily accepted by patients who do not want to advertise their disability. More generally, this poses the general question of how users come to accept and use a technology. This question is debated in the Technology Acceptance Model (TAM) which postulates that two specific perceptions about technology determine one behavioral intention to use a technology: perceived ease of use and perceived usefulness (see, e.g., [50]).

Another important aspect of this work that will progressively need attention is ergonomic which will have to take into account the other potential functional limitations of these patients in addition to low vision (e.g., limitations in mobility, hearing, or agility).

3.2.3. *Biophysical modeling*

Modeling in neuroscience has to cope with several competing objective. On one hand describing the biological realm as close as possible, and, on the other hand, providing tractable equations at least at the descriptive level (simulation, qualitative description) and, when possible, at the mathematical level (i.e., affording a rigorous description). These objectives are rarely achieved simultaneously and most of the time one has to make compromises. In Biovision team we adopt the point of view of physicists: try to capture the phenomenological description of a biophysical mechanism, removing irrelevant details in the description, and try to have a qualitative description of equations behaviour at least at the numerical simulation level, and, when possible, get out analytic results. We do not focus on mathematical proofs, instead insisting on the quality of the model in predicting, and, if possible proposing new experiments. This requires a constant interaction with neuroscientists so as to keep the model on the tracks, warning of too crude approximation, still trying to construct equations from canonical principles [4],[33], [22].

3.2.4. *Methods from theoretical physics*

Biophysical models mainly consist of differential equations (ODEs or PDEs) or integro-differential equations (neural fields). We study them using dynamical systems and bifurcation theory as well as techniques coming from nonlinear physics (amplitude equations, stability analysis, Lyapunov spectrum, correlation analysis, multi-scales methods).

For the study of large scale populations (e.g., when studying population coding) we use methods coming from statistical physics. This branch of physics gave birth to mean-field methods as well statistical methods for large population analysis. We use both of them. Mean-field methods will be applied for large scale activity in the retina and in the cortex [7], [11],[15].

For the study of retina population coding we use the so-called Gibbs distribution, initially introduced by Boltzmann and Gibbs. This concept includes, but *is not limited to*, maximum entropy models [60] used by numerous authors in the context of the retina (see, e.g., [73], [75], [57], [56], [78]). These papers were restricted to a statistical description without memory neither causality: the time correlations between successive times is not considered. A paradigmatic example of this is the Ising model, used to describe the retinal activity in, e.g., [73], [75]. However, maximum entropy extends to spatio-temporal correlations as we have shown in, e.g., [13], [5].

More generally, while maximum entropy models rely heavily on the questionable assumption of stationarity, the concept of Gibbs distribution does not need this hypothesis. Beside, it allows to handle models with large memory; it also provides a framework to model anticipation [16]. It includes as well existing models to explain retina statistics such as the Generalized Linear Model (GLM) [44].

CAMIN Team

3. Research Program

3.1. Exploration and understanding of the origins and control of movement

One of CAMIN's areas of expertise is **motion measurement, observation and modeling** in the context of **sensorimotor deficiencies**. The team has the capacity to design advanced protocols to explore motor control mechanisms in more or less invasive conditions in both animal and human.

Human movement can be assessed by several noninvasive means, from motion observation (MOCAP, IMU) to electrophysiological measurements (afferent ENG, EMG, see below). Our general approach is to develop solutions that are realistic in terms of clinical or home use by clinical staff and/or patients for diagnosis and assessment purposes. In doing so, we try to gain a better understanding of motor control mechanisms, including deficient ones, which in turn will give us greater insight into the basics of human motor control. Our ultimate goal is to optimally match a neuroprosthesis to the targeted sensorimotor deficiency.

The team is involved in research projects including:

- **Peripheral nervous system (PNS) exploration, modeling and electrophysiology techniques**
Electroneurography (ENG) and electromyography (EMG) signals inform about neural and muscular activities. The team investigates both natural and evoked ENG/EMG through advanced and dedicated signal processing methods. Evoked responses to ES are very precious information for understanding neurophysiological mechanisms, as both the input (ES) and the output (evoked EMG/ENG) are controlled. CAMIN has the expertise to perform animal experiments (rabbits, rats, earthworms and big animals with partners), design hardware and software setups to stimulate and record in harsh conditions, process signals, analyze results and develop models of the observed mechanisms. Experimental surgery is mandatory in our research prior to invasive interventions in humans. It allows us to validate our protocols from theoretical, practical and technical aspects.
- **Central nervous system (CNS) exploration**
Stimulating the CNS directly instead of nerves allows activation of the neural networks responsible for generating functions. Once again, if selectivity is achieved the number of implanted electrodes and cables would be reduced, as would the energy demand. We have investigated **spinal electrical stimulation** in animals (pigs) for urinary track and lower limb function management. This work is very important in terms of both future applications and the increase in knowledge about spinal circuitry. The challenges are technical, experimental and theoretical, and the preliminary results have enabled us to test some selectivity modalities through matrix electrode stimulation. This research area will be further intensified in the future as one of ways to improve neuroprosthetic solutions. We intend to gain a better understanding of the electrophysiological effects of DES through electroencephalographic (EEG) and electrocorticographic (ECoG) recordings in order to optimize anatomo-functional brain mapping, better understand brain dynamics and plasticity, and improve surgical planning, rehabilitation, and the quality of life of patients.
- **Muscle models and fatigue exploration**
Muscle fatigue is one of the major limitations in all FES studies. Simply, the muscle torque varies over time even when the same stimulation pattern is applied. As there is also muscle recovery when there is a rest between stimulations, modeling the fatigue is almost an impossible task. Therefore, it is essential to monitor the muscle state and assess the expected muscle response by FES to improve the current FES system in the direction of greater adaptive force/torque control in the presence of muscle fatigue.
- **Movement interpretation**

We intend to develop ambulatory solutions to allow ecological observation. We have extensively investigated the possibility of using inertial measurement units (IMUs) within body area networks to observe movement and assess posture and gait variables. We have also proposed extracting gait parameters like stride length and foot-ground clearance for evaluation and diagnosis purposes.

3.2. Movement assistance and/or restoration

The challenges in movement restoration are: (i) improving nerve/muscle stimulation modalities and efficiency and (ii) global management of the function that is being restored in interaction with the rest of the body under voluntary control. For this, both local (muscle) and global (function) controls have to be considered. Online modulation of ES parameters in the context of lower limb functional assistance requires the availability of information about the ongoing movement. Different levels of complexity can be considered, going from simple open-loop to complex control laws (figure 2).

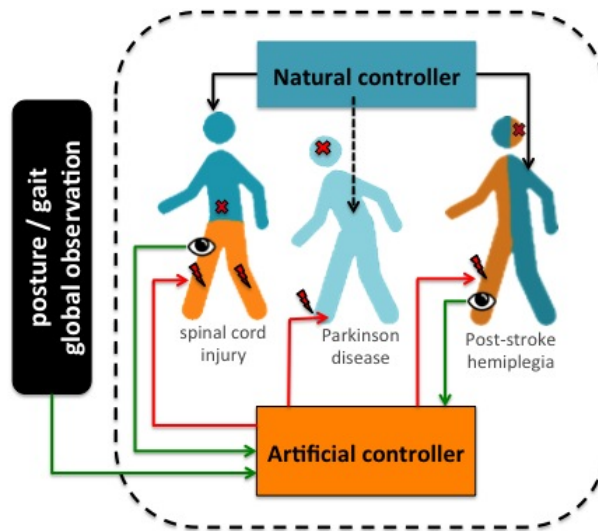


Figure 2. FES assistance should take into account the coexistence of artificial and natural controllers. Artificial controllers should integrate both global (posture/gait) and local (limb/joint) observations.

Real-time adaptation of the stimulation patterns is an important challenge in most of the clinical applications we consider. The modulation of ES parameters in the presence of fatigue or to adapt to context needs for adaptive controllers processing information on movement execution and environmental changes. A minimum number of sensors with minimal impact on patient motion is necessary.

CASTOR Project-Team

3. Research Program

3.1. Plasma Physics

Participants: Jacques Blum, Cédric Boulbe, Blaise Faugeras, Hervé Guillard, Holger Heumann, Sebastian Minjeaud, Boniface Nkonga, Richard Pasquetti, Afeintou Sangam.

The main research topics are:

1. Modelling and analysis
 - Fluid closure in plasma
 - Turbulence
 - Plasma anisotropy type instabilities
 - Free boundary equilibrium (FBE)
 - Coupling FBE – Transport
2. Numerical methods and simulations
 - High order methods
 - Curvilinear coordinate systems
 - Equilibrium simulation
 - Pressure correction scheme
 - Anisotropy
 - Solving methods and parallelism
3. Identification and control
 - Inverse problem: Equilibrium reconstruction
 - Open loop control
4. Applications
 - MHD instabilities : Edge-Localized Modes (ELMs)
 - Edge plasma turbulence
 - Optimization of scenarii

COFFEE Project-Team

3. Research Program

3.1. Research Program

Mathematical modeling and computer simulation are among the main research tools for environmental management, risks evaluation and sustainable development policy. Many aspects of the computer codes as well as the PDEs systems on which these codes are based can be considered as questionable regarding the established standards of applied mathematical modeling and numerical analysis. This is due to the intricate multiscale nature and tremendous complexity of those phenomena that require to set up new and appropriate tools. Our research group aims to contribute to bridging the gap by developing advanced abstract mathematical models as well as related computational techniques.

The scientific basis of the proposal is two-fold. On the one hand, the project is “technically-driven”: it has a strong content of mathematical analysis and design of general methodology tools. On the other hand, the project is also “application-driven”: we have identified a set of relevant problems motivated by environmental issues, which share, sometimes in a unexpected fashion, many common features. The proposal is precisely based on the conviction that these subjects can mutually cross-fertilize and that they will both be a source of general technical developments, and a relevant way to demonstrate the skills of the methods we wish to design.

To be more specific:

- We consider evolution problems describing highly heterogeneous flows (with different phases or with high density ratio). In turn, we are led to deal with non linear systems of PDEs of convection and/or convection–diffusion type.
- The nature of the coupling between the equations can be two-fold, which leads to different difficulties, both in terms of analysis and conception of numerical methods. For instance, the system can couple several equations of different types (elliptic/parabolic, parabolic/hyperbolic, parabolic or elliptic with algebraic constraints, parabolic with degenerate coefficients....). Furthermore, the unknowns can depend on different sets of variables, a typical example being the fluid/kinetic models for particulate flows. In turn, the simulation cannot use a single numerical approach to treat all the equations. Instead, hybrid methods have to be designed which raise the question of fitting them in an appropriate way, both in terms of consistency of the discretization and in terms of stability of the whole computation. For the problems under consideration, the coupling can also arise through interface conditions. It naturally occurs when the physical conditions are highly different in subdomains of the physical domain in which the flows takes place. Hence interface conditions are intended to describe the exchange (of mass, energy...) between the domains. Again it gives rise to rather unexplored mathematical questions, and for numerics it yields the question of defining a suitable matching at the discrete level, that is requested to preserve the properties of the continuous model.
- By nature the problems we wish to consider involve many different scales (of time or length basically). It raises two families of mathematical questions. In terms of numerical schemes, the multiscale feature induces the presence of stiff terms within the equations, which naturally leads to stability issues. A clear understanding of scale separation helps in designing efficient methods, based on suitable splitting techniques for instance. On the other hand asymptotic arguments can be used to derive hierarchy of models and to identify physical regimes in which a reduced set of equations can be used.

We can distinguish the following fields of expertise

- Numerical Analysis: Finite Volume Schemes, Well-Balanced and Asymptotic-Preserving Methods
 - Finite Volume Schemes for Diffusion Equations
 - Finite Volume Schemes for Conservation Laws
 - Well-Balanced and Asymptotic-Preserving Methods
- Modeling and Analysis of PDEs
 - Kinetic equations and hyperbolic systems
 - PDEs in random media
 - Interface problems

LEMON Team

3. Research Program

3.1. State of the Art

3.1.1. Shallow Water Models

Shallow Water (SW) wave dynamics and dissipation represent an important research field. This is because shallow water flows are the most common flows in geophysics. In shallow water regions, dispersive effects (non-hydrostatic pressure effects related to strong curvature in the flow streamlines) can become significant and affect wave transformations. The shoaling of the wave (the “steepening” that happens before the breaking) cannot be described with the usual Saint-Venant equations. To model such various evolutions, one has to use more sophisticated models (Boussinesq, Green-Naghdi...). Nowadays, the classical Saint-Venant equations can be solved numerically in an accurate way, allowing the generation of bores and the shoreline motion to be handled, using recent finite-volume or discontinuous-Galerkin schemes. In contrast, very few advanced works regarding the derivation and modern numerical solution of dispersive equations [23], [27], [56] are available in one dimensions, let alone in the multidimensional case. We can refer to [55], [30] for some linear dispersive equations, treated with finite-element methods, or to [27] for the first use of advanced high-order compact finite-volume methods for the Serre equations. Recent work undertaken during the ANR MathOCEAN [23] lead to some new 1D fully nonlinear and weakly dispersive models (Green-Naghdi like models) that allow to accurately handle the nonlinear waves transformations. High order accuracy numerical methods (based on a second-order splitting strategy) have been developed and implemented, raising a new and promising 1D numerical model. However, there is still a lack of new development regarding the multidimensional case.

In shallow water regions, depending on the complex balance between non-linear effects, dispersive effects and energy dissipation due to wave breaking, wave fronts can evolve into a large range of bore types, from purely breaking to purely undular bore. Boussinesq or Green-Naghdi models can handle these phenomena [21]. However, these models neglect the wave overturning and the associated dissipation, and the dispersive terms are not justified in the vicinity of the singularity. Previous numerical studies concerning bore dynamics using depth-averaged models have been devoted to either purely broken bores using NSW models [24], or undular bores using Boussinesq-type models [34]. Let us also mention [32] for tsunami modeling and [31], [43] for the dam-break problem. A model able to reproduce the various bore shapes, as well as the transition from one type of bore to another, is required. A first step has been made with the one-dimensional code [23], [53]. The SWASH project led by Zijlema at Delft [56] addresses the same issues.

3.1.2. Open boundary conditions and coupling algorithms

For every model set in a bounded domain, there is a need to consider boundary conditions. When the boundaries correspond to a modeling choice rather than to a physical reality, the corresponding boundary conditions should not create spurious oscillations or other unphysical behaviour at the artificial boundary. Such conditions are called **open boundary conditions** (OBC). They have been widely studied by applied mathematicians since the pioneering work of [33] on transparent boundary conditions. Deep studies of these operators have been performed in the case of linear equations, [38], [22], [50]. Unfortunately, in the case of geophysical fluid dynamics, this theory leads to nonlocal conditions (even in linear cases) that are not usable in numerical models. Most of current models (including high quality operational ones) modestly use a *no flux* condition (namely an homogeneous Neumann boundary condition) when a free boundary condition is required. But in many cases, Neumann homogeneous conditions are a very poor approximation of the exact transparent conditions. Hence the need to build higher order approximations of these conditions that remain numerically tractable.

Numerous physical processes are involved in coastal modeling, each of them depending on others (surface winds for coastal oceanography, sea currents for sandbars dynamics, etc.). Connecting two (or more) model solutions at their interface is a difficult task, that is often addressed in a simplified way from the mathematical viewpoint: this can be viewed as the one and only iteration of an iterative process. This results with a low quality coupled system, which could be improved either with additional iterations, and/or thanks to the improvement of interface boundary conditions and the use of OBC (see above). Promising results have been obtained in the framework of **ocean-atmosphere coupling** (in a simplified modeling context) in [44], where the use of advanced coupling techniques (based on domain decomposition algorithm) are introduced.

3.1.3. A need for upscaled shallow water models.

The mathematical modeling of **fluid-biology** coupled systems in lagoon ecosystems requires one or several water models. It is of course not necessary (and not numerically feasible) to use accurate non-hydrostatic turbulent models to force the biological processes over very long periods of time. There is a compromise to be reached between accurate (but untractable) fluid models such as the Navier-Stokes equations and simple (but imprecise) models such as [35].

In urbanized coastal zones, upscaling is also a key issue. This stems not only from the multi-scale aspects dealt with in the previous subsection, but also from modeling efficiency considerations.

The typical size of the relevant hydraulic feature in an urban area is between 0.1 m and 1.0 m, while the size of an urban area usually ranges from 10^3 m to 10^4 m. Refined flow computations (e.g. in simulating the impact of a tsunami) over entire coastal conurbations using a 2D horizontal model thus require 10^6 to 10^9 elements. From an engineering perspective, this makes both the CPU and man-supervised mesh design efforts unaffordable in the present state of technology.

Upscaling provides an answer to this problem by allowing macroscopic equations to be derived from the small-scale governing equations. The powerful, multiple scale expansion-based homogenization technique [20], [19], [49] has been applied successfully to flow and transport upscaling in porous media, but its use is subordinated to the stringent assumptions of (i) the existence of a Representative Elementary Volume (REV), (ii) the scale separation principle, and (iii) the process is not purely hyperbolic at the microscopic scale, otherwise precluding the study of transient solutions [20]. Unfortunately, the REV has been shown recently not to exist in urban areas [37]. Besides, the scale separation principle is violated in the case of sharp transients (such as tsunami waves) impacting urban areas because the typical wavelength is of the same order of magnitude as the microscopic detail (the street/block size). Moreover, 2D shallow water equations are essentially hyperbolic, thus violating the third assumption.

These hurdles are overcome by averaging approaches. Single porosity-based, macroscopic shallow water models have been proposed [29], [36], [39] and applied successfully to urban flood modeling scale experiments [36], [45], [52]. They allow the CPU time to be divided by 10 to 100 compared to classical 2D shallow water models. Recent extensions of these models have been proposed in the form of integral porosity [51] and multiple porosity [37] shallow water models.

3.2. Scientific Objectives

Our main challenge is: build and couple elementary models in coastal areas to improve their capacity to simulate complex dynamics. This challenge consists of three principal scientific objectives. First of all, each of the elementary models has to be consistently developed (regardless of boundary conditions and interactions with other processes). Then open boundary conditions (for the simulation of physical processes in bounded domains) and links between the models (interface conditions) have to be identified and formalized. Finally, models and boundary conditions (*i.e.* coupled systems) should be proposed, analyzed and implemented in a common platform.

3.2.1. Single process models and boundary conditions

The time-evolution of a water flow in a three-dimensional computational domain is classically modeled by Navier-Stokes equations for incompressible fluids. Depending on the physical description of the considered domain, these equations can be simplified or enriched. Consequently, there are **numerous water dynamics models** that are derived from the original Navier-Stokes equations, such as primitive equations, shallow water equations (see [28]), Boussinesq-type dispersive models [21]), etc. The aforementioned models have **very different mathematical natures**: hyperbolic vs parabolic, hydrostatic vs non-hydrostatic, inviscid vs viscous, etc. They all carry nonlinearities that make their mathematical study (existence, uniqueness and regularity of weak and/or strong solutions) highly challenging (not to speak about the \$1M Clay competition for the 3D Navier Stokes equations, which may remain open for some time).

The objective is to focus on the mathematical and numerical modeling of models adapted to **nearshore dynamics**, accounting for complicated wave processes. There exists a large range of models, from the shallow water equations (eventually weakly dispersive) to some fully dispersive deeper models. All these models can be obtained from a suitable asymptotic analysis of the water wave equations (Zakharov formulation) and if the theoretical study of these equations has been recently investigated [42], there is still some serious numerical challenges. So we plan to focus on the derivation and implementation of robust and high order discretization methods for suitable two dimensional models, including enhanced fully nonlinear dispersive models and fully dispersive models, like the Matsuno-generalized approach proposed in [41]. Another objective is to study the shallow water dispersive models without any irrotational flow assumption. Such a study would be of great interest for the study of nearshore circulation (wave induced rip currents).

For obvious physical and/or computational reasons, our models are set in bounded domains. Two types of boundaries are considered: physical and mathematical. Physical boundaries are materialized by an existing interface (atmosphere/ocean, ocean/sand, shoreline, etc.) whereas mathematical boundaries appear with the truncation of the domain of interest. In the latter case, **open boundary conditions** are mandatory in order not to create spurious reflexions at the boundaries. Such boundary conditions being nonlocal and impossible to use in practice, we shall look for approximations. We shall obtain them thanks to the asymptotic analysis of the (pseudo-differential) boundary operators with respect to small parameters (viscosity, domain aspect ratio, Rossby number, etc.). Naturally, we **will seek the boundary conditions leading to the best compromise** between mathematical well-posedness and physical consistency. This will make extensive use of the mathematical theory of **absorbing operators** and their approximations [33].

3.2.2. Coupled systems

The Green-Naghdi equations provide a correct description of the waves up to the breaking point while the Saint-Venant equations are more suitable for the description of the surf zone (i.e. after the breaking). Therefore, the challenge here is first to **design a coupling strategy** between these two systems of equations, first in a simplified one-dimensional case, then to the two-dimensional case both on cartesian and unstructured grids. High order accuracy should be achieved through the use of flexible Discontinuous-Galerkin methods.

Additionally, we will couple our weakly dispersive shallow water models to other fully dispersive deeper water models. We plan to mathematically analyze the coupling between these models. In a first step, we have to understand well the mixed problem (initial and boundary conditions) for these systems. In a second step, these new mathematical development have to be embedded within a numerically efficient strong coupling approach. The deep water model should be fully dispersive (solved using spectral methods, for instance) and the shallow-water model will be, in a first approach, the Saint-Venant equations. Then, when the 2D extension of the currently developed Green-Naghdi numerical code will be available, the improved coupling with a weakly dispersive shallow water model should be considered.

In the context of Schwarz relaxation methods, usual techniques can be seen as the first iteration (not converged) of an iterative algorithm. Thanks to the work performed on efficient boundary conditions, we shall **improve the quality of current coupling algorithms**, allowing for qualitatively satisfying solutions **with a reduced computational cost** (small number of iterations).

We are also willing to explore the role of geophysical processes on some biological ones. For example, the design of optimal shellfish farms relies on confinement maps and plankton dynamics, which strongly depend on long-time averaged currents. Equations that model the time evolution of species in a coastal ecosystem are relatively simple from a modeling viewpoint: they mainly consist of ODEs, and possibly advection-diffusion equations. The issue we want to tackle is the choice of the fluid model that should be coupled to them, accounting for the important time scales discrepancy between biological (evolution) processes and coastal fluid dynamics. Discrimination criteria between refined models (such as turbulent Navier-Stokes) and cheap ones (see [35]) will be proposed.

Coastal processes evolve at very different time scales: atmosphere (seconds/minutes), ocean (hours), sediment (months/years) and species evolution (years/decades). Their coupling can be seen as a *slow-fast* dynamical system, and a naïve way to couple them would be to pick the smallest time-step and run the two models together: but the computational cost would then be way too large. Consequently **homogenization techniques or other upscaling methods** should be used in order to account for these various time scales at an affordable computational cost. The research objectives are the following:

- So far, the proposed upscaled models have been validated against theoretical results obtained from refined 2D shallow water models and/or very limited data sets from scale model experiments. The various approaches proposed in the literature [25], [26], [29], [36], [37], [39], [45], [51], [52] have not been compared over the same data sets. Part of the research effort will focus on the extensive validation of the models on the basis of scale model experiments. Active cooperation will be sought with a number of national and international Academic partners involved in urban hydraulics (UCL Louvain-la-Neuve, IMFS Strasbourg, Irvine University California) with operational experimental facilities.
- Upscaling of source terms. Two types of source terms play a key role in shallow water models: geometry-induced source terms (arising from the irregular bathymetry) and friction/turbulence-induced energy loss terms. In all the upscaled shallow water models presented so far, only the large scale effects of topographical variations have been upscaled. In the case of wetting/drying phenomena and small depths (e.g. the *Camargue* tidal flats), however, it is foreseen that subgrid-scale topographic variations may play a predominant role. Research on the integration of subgrid-scale topography into macrosopic shallow water models is thus needed. Upscaling of friction/turbulence-induced head loss terms is also a subject for research, with a number of competing approaches available from the literature [36], [37], [51], [54].
- Upscaling of transport processes. The upscaling of surface pollutant transport processes in the urban environment has not been addressed so far in the literature. Free surface flows in urban areas are characterized by strongly variable (in both time and space) flow fields. Dead/swirling zones have been shown to play a predominant role in the upscaling of the flow equations [37], [51]. Their role is expected to be even stronger in the upscaling of contaminant transport. While numerical experiments indicate that the microscopic hydrodynamic time scales are small compared to the macroscopic time scales, theoretical considerations indicate that this may not be the case with scalar transport. Trapping phenomena at the microscopic scale are well-known to be upscaled in the form of fractional dynamics models in the long time limit [40], [47]. The difficulty in the present research is that upscaling is not sought only for the long time limit but also for all time scales. Fractional dynamics will thus probably not suffice to a proper upscaling of the transport equations at all time scales.

3.2.3. Numerical platform

As a long term objective, the team shall create a common architecture for existing codes, and also the future codes developed by the project members, to offer a simplified management of various evolutions and a single and well documented tool for our partners. It will aim to be self-contained including pre and post-processing tools (efficient meshing approaches, GMT and VTK libraries), but must of course also be opened to user's suggestions, and account for existing tools inside and outside Inria. This numerical platform will be dedicated to the simulation of all the phenomena of interest, including flow propagation, sediment evolution, model

coupling on large scales, from deep water to the shoreline, including swell propagation, shoaling, breaking and run-up. This numerical platform clearly aims at becoming a reference software in the community. It should be used to **develop a specific test case** around Montpellier which embeds many processes and their mutual interactions: from the *Camargue* (where the Rhône river flows into the Mediterranean sea) to the *Étang de Thau* (a wide lagoon where shellfishes are plentiful), **all the processes studied in the project occur in a 100km wide region**, including of course the various hydrodynamics regimes (from the deep sea to the shoaling, surf and swash zones) and crucial morphodynamic issues (*e.g.* in the town of Sete).

MATHNEURO Team

3. Research Program

3.1. Neural networks dynamics

The study of neural networks is certainly motivated by the long term goal to understand how brain is working. But, beyond the comprehension of brain or even of simpler neural systems in less evolved animals, there is also the desire to exhibit general mechanisms or principles at work in the nervous system. One possible strategy is to propose mathematical models of neural activity, at different space and time scales, depending on the type of phenomena under consideration. However, beyond the mere proposal of new models, which can rapidly result in a plethora, there is also a need to understand some fundamental keys ruling the behaviour of neural networks, and, from this, to extract new ideas that can be tested in real experiments. Therefore, there is a need to make a thorough analysis of these models. An efficient approach, developed in our team, consists of analysing neural networks as dynamical systems. This allows to address several issues. A first, natural issue is to ask about the (generic) dynamics exhibited by the system when control parameters vary. This naturally leads to analyse the bifurcations [8] [37] occurring in the network and which phenomenological parameters control these bifurcations. Another issue concerns the interplay between neuron dynamics and synaptic network structure.

3.2. Mean-field approaches

Modeling neural activity at scales integrating the effect of thousands of neurons is of central importance for several reasons. First, most imaging techniques are not able to measure individual neuron activity (microscopic scale), but are instead measuring mesoscopic effects resulting from the activity of several hundreds to several hundreds of thousands of neurons. Second, anatomical data recorded in the cortex reveal the existence of structures, such as the cortical columns, with a diameter of about $50\mu m$ to $1mm$, containing of the order of one hundred to one hundred thousand neurons belonging to a few different species. The description of this collective dynamics requires models which are different from individual neurons models. In particular, when the number of neurons is large enough averaging effects appear, and the collective dynamics is well described by an effective mean-field, summarizing the effect of the interactions of a neuron with the other neurons, and depending on a few effective control parameters. This vision, inherited from statistical physics requires that the space scale be large enough to include a large number of microscopic components (here neurons) and small enough so that the region considered is homogeneous.

Our group is developing mathematical and numerical methods allowing on one hand to produce dynamic mean-field equations [1] [36] from the physiological characteristics of neural structure (neurons type, synapse type and anatomical connectivity between neurons populations), and on the other so simulate these equations.

3.3. Neural fields

Neural fields are a phenomenological way of describing the activity of population of neurons by delay integro-differential equations. This continuous approximation turns out to be very useful to model large brain areas such as those involved in visual perception. The mathematical properties of these equations and their solutions are still imperfectly known, in particular in the presence of delays, different time scales and of noise.

Our group is developing mathematical and numerical methods for analysing these equations. These methods are based upon techniques from mathematical functional analysis, bifurcation theory [9], equivariant bifurcation analysis, delay equations, and stochastic partial differential equations. We have been able to characterize the solutions of these neural fields equations and their bifurcations, apply and expand the theory to account for such perceptual phenomena as edge, texture [3], and motion perception. We have also developed a theory of the delayed neural fields equations, in particular in the case of constant delays and propagation delays that must be taken into account when attempting to model large size cortical areas [38]. This theory is based on center manifold and normal forms ideas.

3.4. Slow-Fast Dynamics in Neuronal Models

Neuronal rhythms typically display many different timescales, therefore it is important to incorporate this slow-fast aspect in models. We are interested in this modeling paradigm where slow-fast point models (using Ordinary Differential Equations) are investigated in terms of their bifurcation structure and the patterns of oscillatory solutions that they can produce. To insight into the dynamics of such systems, we use a mix of theoretical techniques — such as geometric desingularisation and centre manifold reduction [35] — and numerical methods such as pseudo-arclength continuation [32]. We are interested in families of complex oscillations generated by both mathematical and biophysical models of neurons. In particular, so-called *mixed-mode oscillations (MMOs)* [30], [34]), which represent an alternation between subthreshold and spiking behaviour, and *bursting oscillations* [31], [33], also corresponding to experimentally observed behaviour [29].

Selected publications on this topic: [lien](#).

3.5. Synaptic Plasticity

Neural networks show amazing abilities to evolve and adapt, and to store and process information. These capabilities are mainly conditioned by plasticity mechanisms, and especially synaptic plasticity, inducing a mutual coupling between network structure and neuron dynamics. Synaptic plasticity occurs at many levels of organization and time scales in the nervous system [28]. It is of course involved in memory and learning mechanisms, but it also alters excitability of brain areas and regulates behavioral states (e.g. transition between sleep and wakeful activity). Therefore, understanding the effects of synaptic plasticity on neurons dynamics is a crucial challenge.

Our group is developing mathematical and numerical methods to analyse this mutual interaction. On the one hand, we have shown that plasticity mechanisms, Hebbian-like or STDP, have strong effects on neuron dynamics complexity, such as dynamics complexity reduction, and spike statistics

3.6. Visual Neuroscience

Our group focuses on the visual system to understand how information is encoded and processed resulting in visual percepts. To do so, we propose functional models of the visual system using a variety of mathematical formalisms, depending on the scale at which models are built, such as spiking neural networks or neural fields. So far, our efforts have been focused on the study of retinal processing, edge and texture perception, motion integration at the level of V1 and MT cortical areas.

MORPHEME Project-Team

3. Research Program

3.1. Research Program

The recent advent of an increasing number of new microscopy techniques giving access to high throughput screenings and micro or nano-metric resolutions provides a means for quantitative imaging of biological structures and phenomena. To conduct quantitative biological studies based on these new data, it is necessary to develop non-standard specific tools. This requires using a multi-disciplinary approach. We need biologists to define experiment protocols and interpret the results, but also physicists to model the sensors, computer scientists to develop algorithms and mathematicians to model the resulting information. These different expertises are combined within the Morpheme team. This generates a fecund frame for exchanging expertise, knowledge, leading to an optimal framework for the different tasks (imaging, image analysis, classification, modeling). We thus aim at providing adapted and robust tools required to describe, explain and model fundamental phenomena underlying the morphogenesis of cellular and supra-cellular biological structures. Combining experimental manipulations, *in vivo* imaging, image processing and computational modeling, we plan to provide methods for the quantitative analysis of the morphological changes that occur during development. This is of key importance as the morphology and topology of mesoscopic structures govern organ and cell function. Alterations in the genetic programs underlying cellular morphogenesis have been linked to a range of pathologies.

Biological questions we will focus on include:

1. what are the parameters and the factors controlling the establishment of ramified structures? (Are they really organize to ensure maximal coverage? How are genetic and physical constraints limiting their morphology?),
2. how are newly generated cells incorporated into reorganizing tissues during development? (is the relative position of cells governed by the lineage they belong to?)

Our goal is to characterize different populations or development conditions based on the shape of cellular and supra-cellular structures, e.g. micro-vascular networks, dendrite/axon networks, tissues from 2D, 2D+t, 3D or 3D+t images (obtained with confocal microscopy, video-microscopy, photon-microscopy or micro-tomography). We plan to extract shapes or quantitative parameters to characterize the morphometric properties of different samples. On the one hand, we will propose numerical and biological models explaining the temporal evolution of the sample, and on the other hand, we will statistically analyze shapes and complex structures to identify relevant markers for classification purposes. This should contribute to a better understanding of the development of normal tissues but also to a characterization at the supra-cellular scale of different pathologies such as Alzheimer, cancer, diabetes, or the Fragile X Syndrome. In this multidisciplinary context, several challenges have to be faced. The expertise of biologists concerning sample generation, as well as optimization of experimental protocols and imaging conditions, is of course crucial. However, the imaging protocols optimized for a qualitative analysis may be sub-optimal for quantitative biology. Second, sample imaging is only a first step, as we need to extract quantitative information. Achieving quantitative imaging remains an open issue in biology, and requires close interactions between biologists, computer scientists and applied mathematicians. On the one hand, experimental and imaging protocols should integrate constraints from the downstream computer-assisted analysis, yielding to a trade-off between qualitative optimized and quantitative optimized protocols. On the other hand, computer analysis should integrate constraints specific to the biological problem, from acquisition to quantitative information extraction. There is therefore a need of specificity for embedding precise biological information for a given task. Besides, a level of generality is also desirable for addressing data from different teams acquired with different protocols and/or sensors. The mathematical modeling of the physics of the acquisition system will yield higher performance reconstruction/restoration algorithms in terms of accuracy. Therefore, physicists and computer scientists have to work together. Quantitative information extraction also has to deal with both the complexity of the structures of interest (e.g., very

dense network, small structure detection in a volume, multiscale behavior, ...) and the unavoidable defects of in vivo imaging (artifacts, missing data, ...). Incorporating biological expertise in model-based segmentation methods provides the required specificity while robustness gained from a methodological analysis increases the generality. Finally, beyond image processing, we aim at quantifying and then statistically analyzing shapes and complex structures (e.g., neuronal or vascular networks), static or in evolution, taking into account variability. In this context, learning methods will be developed for determining (dis)similarity measures between two samples or for determining directly a classification rule using discriminative models, generative models, or hybrid models. Besides, some metrics for comparing, classifying and characterizing objects under study are necessary. We will construct such metrics for biological structures such as neuronal or vascular networks. Attention will be paid to computational cost and scalability of the developed algorithms: biological experiments generally yield huge data sets resulting from high throughput screenings. The research of Morpheme will be developed along the following axes:

- **Imaging:** this includes i) definition of the studied populations (experimental conditions) and preparation of samples, ii) definition of relevant quantitative characteristics and optimized acquisition protocol (staining, imaging, ...) for the specific biological question, and iii) reconstruction/restoration of native data to improve the image readability and interpretation.
- **Feature extraction:** this consists in detecting and delineating the biological structures of interest from images. Embedding biological properties in the algorithms and models is a key issue. Two main challenges are the variability, both in shape and scale, of biological structures and the huge size of data sets. Following features along time will allow to address morphogenesis and structure development.
- **Classification/Interpretation:** considering a database of images containing different populations, we can infer the parameters associated with a given model on each dataset from which the biological structure under study has been extracted. We plan to define classification schemes for characterizing the different populations based either on the model parameters, or on some specific metric between the extracted structures.
- **Modeling:** two aspects will be considered. This first one consists in modeling biological phenomena such as axon growing or network topology in different contexts. One main advantage of our team is the possibility to use the image information for calibrating and/or validating the biological models. Calibration induces parameter inference as a main challenge. The second aspect consists in using a prior based on biological properties for extracting relevant information from images. Here again, combining biology and computer science expertise is a key point.

VIRTUAL PLANTS Project-Team

3. Research Program

3.1. Analysis of structures resulting from meristem activity

To analyze plant growth and structure, we focus mainly on methods for analyzing sequences and tree-structured data. These methods range from algorithms for computing distance between sequences or tree-structured data to statistical models.

- *Combinatorial approaches*: plant structures exhibit complex branching organizations of their organs like internodes, leaves, shoots, axes, branches, etc. These structures can be analyzed with combinatorial methods in order to compare them or to reveal particular types of organization. We investigate a family of techniques to quantify distances between branching systems based on non-linear structural alignment (similar to edit-operation methods used for sequence comparison). Based on these techniques, we study the notion of (topology-based) self-similarity of branching structures in order to define a notion of degree of redundancy for any tree structure and to quantify in this way botanical notions, such as the physiological states of a meristem, fundamental to the description of plant morphogenesis.
- *Statistical modeling*: We investigate different categories of statistical models corresponding to different types of structures.
 - Longitudinal data corresponding to plant growth follow up: the statistical models of interest are equilibrium renewal processes and generalized linear mixed models for longitudinal count data.
 - Repeated patterns within sequences or trees: the statistical models of interest are mainly (hidden) variable-order Markov chains. Hidden variable-order Markov chains were in particular applied to characterize permutation patterns in phyllotaxis and the alternation between flowering and vegetative growth units along sympodial tree axes.
 - Homogeneous zones (or change points) within sequences or trees: most of the statistical models of interest are hidden Markovian models (hidden semi-Markov chains, semi-Markov switching linear mixed models and semi-Markov switching generalized linear models for sequences and different families of hidden Markov tree models). A complementary approach consists in applying multiple change-point models. The branching structure of a parent shoot is often organized as a succession of branching zones while the succession of shoot at the more macroscopic scale exhibit roughly stationary phases separated by marked change points.

We investigate both estimation methods and diagnostic tools for these different categories of models. In particular we focus on diagnostic tools for latent structure models (e.g. hidden Markovian models or multiple change-point models) that consist in exploring the latent structure space.

- *A new generation of morphogenesis models*: Designing morphogenesis models of the plant development at the macroscopic scales is a challenging problem. As opposed to modeling approaches that attempt to describe plant development on the basis of the integration of purely mechanistic models of various plant functions, we intend to design models that tightly couple mechanistic and empirical sub-models that are elaborated in our plant architecture analysis approach. Empirical models are used as a powerful complementary source of knowledge in places where knowledge about mechanistic processes is lacking or weak. We chose to implement such integrated models in a programming language dedicated to dynamical systems with dynamical structure $(DS)^2$, such as L-systems or MGS. This type of language plays the role of an integration framework for sub-models of heterogeneous nature.

3.2. Meristem functioning and development

In this second scientific axis, we develop models of meristem growth at tissue level in order to integrate various sources of knowledge and to analyze their dynamic and complex spatial interaction. To carry out this integration, we need to develop a complete methodological approach containing:

- algorithms for the automatized segmentation in 3D, and cell lineage tracking throughout time, for images coming from confocal microscopy,
- design of high-level routines and user interfaces to distribute these image analysis tools to the scientific community,
- tools for structural and statistical analysis of 3D meristem structure (spatial statistics, multiscale geometric and topological analysis),
- physical models of cells interactions based on spring-mass systems or on tensorial mechanics at the level of cells,
- models of biochemical networks of hormonal and gene driven regulation, at the cellular and tissue level, using continuous and discrete formalisms,
- and models of cell development taking into account the effects of growth and cell divisions on the two previous classes of models.

COATI Project-Team

3. Research Program

3.1. Research Program

Members of COATI have a strong expertise in the design and management of wired and wireless backbone, backhaul, broadband, and complex networks. On the one hand, we cope with specific problems such as energy efficiency in backhaul and backbone networks, routing reconfiguration in connection oriented networks (MPLS, WDM), traffic aggregation in SONET networks, compact routing in large-scale networks, survivability to single and multiple failures, etc. These specific problems often come from questions of our industrial partners. On the other hand, we study fundamental problems mainly related to routing and reliability that appear in many networks (not restricted to our main fields of applications) and that have been widely studied in the past. However, previous solutions do not take into account the constraints of current networks/traffic such as their huge size and their dynamics. COATI thus puts a significant research effort in the following directions:

- **Energy efficiency and Software-Defined Networks (SDN)** at both the design and management levels. More precisely, we plan to study the deployment of energy-efficient routing algorithm within SDN. We developed new algorithms in order to take into account the new constraints of SDN equipments and we evaluate their performance by simulation and by experimentation on a fat-tree architecture.
- **Larger networks:** Another challenge one has to face is the increase in size of practical instances. It is already difficult, if not impossible, to solve practical instances optimally using existing tools. Therefore, we have to find new ways to solve problems using reduction and decomposition methods, characterization of polynomial instances (which are surprisingly often the practical ones), or algorithms with acceptable practical performances.
- **Stochastic behaviors:** Larger topologies mean frequent changes due to traffic and radio fluctuations, failures, maintenance operations, growth, routing policy changes, etc. We aim at including these stochastic behaviors in our combinatorial optimization process to handle the dynamics of the system and to obtain robust designs of networks.

DIANA Project-Team

3. Research Program

3.1. Service Transparency

Transparency is to provide network users and application developers with reliable information about the current or predicted quality of their communication services, and about potential leakages of personal information, or of other information related to societal interests of the user as a “connected citizen” (e.g. possible violation of network neutrality, opinion manipulation). Service transparency therefore means to provide information meaningful to users and application developers, such as quality of experience, privacy leakages, or opinion manipulation, etc. rather than network-level metrics such as available bandwidth, loss rate, delay or jitter.

The Internet is built around a best effort routing service that does not provide any guarantee to end users in terms of quality of service (QoS). The simplicity of the Internet routing service is at the root of its huge success. Unfortunately, a simple service means unpredicted quality at the access. Even though a considerable effort is done by operators and content providers to optimise the Internet content delivery chain, mainly by over-provisioning and sophisticated engineering techniques, service degradation is still part of the Internet. The proliferation of wireless and mobile access technologies, and the versatile nature of Internet traffic, make end users quality of experience (QoE) forecast even harder. As a matter of fact, the Internet is missing a dedicated measurement plane that informs the end users on the quality they obtain and in case of substantial service degradation, on the origin of this degradation. Current state of the art activities are devoted to building a distributed measurement infrastructure to perform active, passive and hybrid measurements in the wired Internet. However, the problem is exacerbated with modern terminals such as smartphones or tablets that do not facilitate the task for end users (they even make it harder) as they focus on simplifying the interface and limiting the control on the network, whereas the Internet behind is still the same in terms of the quality it provides. Interestingly, this same observation explains the existing difficulty to detect and prevent privacy leaks. We argue that the lack of transparency for diagnosing QoE and for detecting privacy leaks have the same root causes and can be solved using common primitives. For instance, in both cases, it is important to be able to link data packets to an application. Indeed, as the network can only access data packets, there must be a way to bind these packets to an application (to understand users QoE for this application or to associate a privacy leak to an application). This is however a complex task as the traffic might be obfuscated or encrypted. Our objectives in the research direction are the following:

- Design and develop measurement tools providing transparency, in spite of current complexity
- Deploy those measurement tools at the Internet’s edge and make them useful for end users
- Propose measurements plane as an overlay or by exploiting in-network functionalities
- Adapt measurements techniques to network architectural change
- Provide measurements as native functionality in future network architecture

3.2. Open network architecture

We are surrounded by personal content of all types: photos, videos, documents, etc. The volume of such content is increasing at a fast rate, and at the same time, the spread of such content among all our connected devices (mobiles, storage devices, set-top boxes, etc) is also increasing. All this complicates the control of personal content by the user both in terms of access and sharing with other users. The access of the personal content in a seamless way independently of its location is a key challenge for the future of networks. Proprietary solutions exist, but apart from fully depending on one of them, there is no standard plane in the Internet for a seamless access to personal content. Therefore, providing network architectural support to design and develop content access and sharing mechanisms is crucial to allow users control their own data over heterogeneous underlying network or cloud services.

On the other hand, privacy is a growing concern for states, administrations, and companies. Indeed, for instance the French CNIL (entity in charge of citizens privacy in computer systems) puts privacy at the core of its activities by defining rules on any stored and collected private data. Also, companies start to use privacy preserving solutions as a competitive advantage. Therefore, understanding privacy leaks and preventing them is a problem that can already find support. However, all end-users do not *currently* put privacy as their first concern. Indeed, in face of two services with one of higher quality, they usually prefer the highest quality one whatever the privacy implication. This was, for instance, the case concerning the Web search service of Google that is more accurate but less privacy preserving than Bing. This is also the case for cloud services such as iCloud or Dropbox that are much more convenient than open source solutions, but very bad in terms of privacy. Therefore, to reach end-users, any privacy preserving solutions must offer a service equivalent to the best existing services.

We consider that it will be highly desirable for Internet users to be able to *easily* move their content from a provider to another and therefore not to depend on a content provider or a social network monopoly. This requires that the network provides built-in architectural support for content networking.

In this research direction, we will define a new *service abstraction layer* (SAL) that could become the new waist of the network architecture with network functionalities below (IP, SDN, cloud) and applications on top. SAL will define different services that are of use to all Internet users for accessing and sharing data (seamless content localisation and retrieval, privacy leakage protection, transparent vertical and horizontal handover, etc.). The biggest challenge here is to cope in the same time with large number of content applications requirements and high underlying networks heterogeneity while still providing efficient applications performance. This requires careful definition of the services primitives and the parameters to be exchanged through the service abstraction layer.

Two concurring factors make the concept behind SAL feasible and relevant today. First, the notion of scalable network virtualization that is a required feature to deploy SAL in real networks today has been discussed recently only. Second, the need for new services abstraction is recent. Indeed, fifteen years ago the Internet for the end-users was mostly the Web. Only eight years ago smartphones came into the picture of the Internet boosting the number of applications with new functionalities and risks. Since a few years, many discussions in the network communities took place around the actual complexity of the Internet and the difficulty to develop applications. Many different approaches have been discussed (such as CCN, SDN) that intend to solve only part of the complexity. SAL takes a broader architectural look at the problem and considers solutions such as CCN as mere use cases. Our objectives in this research direction include the following:

- Identify common key networking services required for content access and sharing
- Detect and prevent privacy leaks for content communication
- Enhance software defined networks for large scale heterogeneous environments
- Design and develop open Content Networking architecture
- Define a service abstraction layer as the thin waist for the future content network architecture
- Test and deploy different applications using SAL primitives on heterogeneous network technologies

3.3. Methodology

We follow an experimental approach that can be described in the following techniques:

- Measurements: the aim is to get a better view of a problem in quantifiable terms. Depending on the field of interest, this may involve large scale distributed systems crawling tools; active probing techniques to infer the status and properties of a complex and non controllable system as the Internet; or even crowdsourcing-based deployments for gathering data on real-users environments or behaviours.
- Experimental evaluation: once a new idea has been designed and implemented, it is of course very desirable to assess and quantify how effective it can be, before being able to deploy it on any realistic scale. This is why a wide range of techniques can be considered for getting early, yet as significant as possible, feedback on a given paradigm or implementation. The spectrum for such techniques span from simulations to real deployments in protected and/or controlled environments.

FOCUS Project-Team

3. Research Program

3.1. Models

The objective of Focus is to develop concepts, techniques, and possibly also tools, that may contribute to the analysis and synthesis of CBUS. Fundamental to these activities is *modeling*. Therefore designing, developing and studying computational models appropriate for CBUS is a central activity of the project. The models are used to formalise and verify important computational properties of the systems, as well as to propose new linguistic constructs.

The models we study are in the process calculi (e.g., the π -calculus) and λ -calculus tradition. Such models, with their emphasis on algebra, well address compositionality—a central property in our approach to problems. Accordingly, the techniques we employ are mainly operational techniques based on notions of behavioural equivalence, and techniques based on algebra, mathematical logics, and type theory.

INDES Project-Team

3. Research Program

3.1. Parallelism, concurrency, and distribution

Concurrency management is at the heart of diffuse programming. Since the execution platforms are highly heterogeneous, many different concurrency principles and models may be involved. Asynchronous concurrency is the basis of shared-memory process handling within multiprocessor or multicore computers, of direct or fifo-based message passing in distributed networks, and of fifo- or interrupt-based event handling in web-based human-machine interaction or sensor handling. Synchronous or quasi-synchronous concurrency is the basis of signal processing, of real-time control, and of safety-critical information acquisition and display. Interfacing existing devices based on these different concurrency principles within HOP or other diffuse programming languages will require better understanding of the underlying concurrency models and of the way they can nicely cooperate, a currently ill-resolved problem.

3.2. Web and functional programming

We are studying new paradigms for programming Web applications that rely on multi-tier functional programming [8]. We have created a Web programming environment named HOP. It relies on a single formalism for programming the server-side and the client-side of the applications as well as for configuring the execution engine.

HOP is a functional language based on the SCHEME programming language. That is, it is a strict functional language, fully polymorphic, supporting side effects, and dynamically type-checked. HOP is implemented as an extension of the BIGLOO compiler that we develop [9]. In the past, we have extensively studied static analyses (type systems and inference, abstract interpretations, as well as classical compiler optimizations) to improve the efficiency of compilation in both space and time.

3.3. Security of diffuse programs

The main goal of our security research is to provide scalable and rigorous language-based techniques that can be integrated into multi-tier compilers to enforce the security of diffuse programs. Research on language-based security has been carried on before in former Inria teams [2], [1]. In particular previous research has focused on controlling information flow to ensure confidentiality.

Typical language-based solutions to these problems are founded on static analysis, logics, provable cryptography, and compilers that generate correct code by construction [6]. Relying on the multi-tier programming language HOP that tames the complexity of writing and analysing secure diffuse applications, we are studying language-based solutions to prominent web security problems such as code injection and cross-site scripting, to name a few.

MAESTRO Project-Team

3. Research Program

3.1. Research Directions

MAESTRO's research directions belong to five main themes motivated by direct applications: network science, wireless networks, network engineering games, green networking and smart grids, content-oriented systems. These directions are very connected: network engineering games find applications in many networking fields, from wireless protocols to applications such as social networks. Green IT studies are often concerned with wireless networks, etc. The study of these applications often raises questions of methodological nature, less close to direct applications; these advances are reported in a separate section.

3.1.1. Network Science

MAESTRO contributes to this new fast growing research subject. "Network Science" or "Complex Network Analysis" aims at understanding the structural properties and the dynamics of a variety of large-scale networks in telecommunications (e.g. the graph of autonomous systems, the Web graph), social science (e.g. community of interest, advertisement, reputation, recommendation systems), bibliometrics (e.g. citations, co-authors), biology (e.g. spread of an epidemic, protein-protein interactions), and physics. It has been observed that the complex networks encountered in these areas share common properties such as power law degree distribution, small average distances, community structure, etc. It also appears that many general questions/applications (e.g. community detection, epidemic spreading, search, anomaly detection) are common in various disciplines which study networks. In particular, we aim at understanding the evolution of complex networks with the help of game theoretical tools in connection with Network Engineering Games, as described below. We design efficient tools for measuring specific properties of large scale complex networks and their dynamics. More specifically, we work on the problem of distributed optimization in large networks where nodes cooperatively solve an optimization problem relying only on local information exchange.

3.1.2. Wireless Networks

The amazing technological advances in wireless devices has led networks to become heterogeneous and very complex. Many research groups worldwide investigate performance evaluation of wireless technologies. MAESTRO's specificity relies on the use of a large variety of analytic tools from applied probability, control theory and distributed optimization to study and improve wireless networks functionalities. We investigate in particular problems of self-organization, channel selection and power control, the association problem and others.

3.1.3. Network Engineering Games

The foundations of *Network Engineering Games* are currently being laid. These are games arising in telecommunications engineering at all the networking layers. This includes considerations from information and communications theory for dealing with the physical and link layers, along with cross layer approaches. MAESTRO's focus is on three areas: *routing games*, *evolutionary games* and *epidemic games*. In routing games we progress on the theory for costs that are not additive over links (such as packet losses or call blocking probabilities). We pursue their research in the stochastic extension of evolutionary game theory, namely the "anonymous sequential games" in which we study the total expected costs and the average cost. Within epidemic games they study epidemics that compete against each other. We apply this to social networks, considering in particular the coupling between various social networks (e.g. propagation strategies that combine Twitter, FaceBook and other social networks).

3.1.4. Green Networking and Smart Grids

The ICT (Information and Communications Technology) sector is becoming one of the main energy consumers worldwide. There is awareness that networks should have a reduced environmental footprint. Our objective is to have a systematically “green” approach when solving optimization problems. The energy cost and the environmental impact should be considered in optimization functions along with traditional performance metrics such as throughput, fairness or delay. We aim at contributing to the design and the analysis of future green networks, in particular those using renewable energy.

Researchers envision that future electricity distribution network will be “smart”, with a large number of small generators (due to an extensive use of renewable energies) and of consumer devices able to adapt their energy needs to a time-varying offer. Generators and devices will be able to locally communicate through the electrical grid itself (or more traditional communication networks), in order to optimize production, transport and use of the energy. This is definitely a new application scenario for MAESTRO, to which we hope to be able to contribute with our expertise on analytic models and performance evaluation.

3.1.5. Content-Oriented Systems

We generally study problems related with the placement and the retrieval of data in communication networks.

We are particularly interested in In-network caching, a widely adopted technique to provide an efficient access to data or resources on a world-wide deployed system while ensuring scalability and availability. For instance, caches are integral components of the Domain Name System, the World Wide Web, Content Distribution Networks, or the recently proposed Information-Centric Network (ICN) architectures. We analyze network of caches, study their optimal placement in the network and optimize data placement in caches/servers.

We also study other aspects related to replication and placement of data: how much to replicate it and on which servers to place it? Finally, we study optimal ways of retrieving the data through prefetching.

3.1.6. Advances in Methodological Tools

MAESTRO has a methodological activity that aims at advancing the state of the art in the tools used for the general performance evaluation and control of systems. We contribute to such fields as perturbation analysis, Markov processes, queueing theory, control theory and game theory. Another objective is to enhance our activity on general-purpose modeling algorithms and software for controlled and uncontrolled stochastic systems.

3.2. Scientific Foundations

The main mathematical tools and formalisms used in MAESTRO include:

- theory of stochastic processes: Markov process, renewal process, branching process, point process, Palm measure, large deviations, mean-field approximation, fluid approximation;
- theory of dynamical discrete-event systems: queues, pathwise and stochastic comparisons, random matrix theory;
- theory of control and scheduling: dynamic programming, Markov decision process, game theory, deterministic and stochastic scheduling; stochastic approximation algorithms;
- theory of singular perturbations.

AYIN Team

3. Research Program

3.1. Geometric and shape modeling

One of the grand challenges of computer vision and image processing is the expression and use of prior geometric information via the construction of appropriate models. For very high resolution imagery, this problem becomes critically important, as the increasing resolution of the data results in the appearance of a great deal of complex geometric structure hitherto invisible. AYIN studies various approaches to the construction of models of geometry and shape.

3.1.1. Stochastic geometry

One of the most promising approaches to the inclusion of this type of information is stochastic geometry, which is an important research direction in the AYIN team. Instead of defining probabilities for different types of image, probabilities are defined for configurations of an indeterminate number of interacting, parameterized objects located in the image. Such probability distributions are called ‘marked point processes’. New models are being developed both for remote sensing applications, and for skin care problems, such as wrinkle and acne detection.

3.1.2. Contours, phase fields, and MRFs with long-range interactions

An alternative approach to shape modeling starts with generic ‘regions’ in the image, and adds constraints in order to model specific shapes and objects. AYIN investigates contour, phase field, and binary field representations of regions, incorporating shape information via highly-structured long-range interactions that constrain the set of high-probability regions to those with specific geometric properties. This class of models can represent infinite-dimensional families of shapes and families with unbounded topology, as well as families consisting of an arbitrary number of object instances, at no extra computational cost. Key sub-problems include the development of models of more complex shapes and shape configurations; the development of models in more than two spatial dimensions; and understanding the equivalences between models in different representations and approaches.

3.1.3. Shapes in time

AYIN is concerned with spectral and spatio-temporal structures. To deal with the latter, the above scene modeling approaches are extended into the time dimension, either by modeling time dependence directly, or, in the field-based approaches, by modeling spacetime structures, or, in the stochastic geometry approach, by including the time t in the mark. An example is a spatio-temporal graph-cut-based method that introduces directed infinite links connecting pixels in successive image frames in order to impose constraints on shape change.

3.2. Image modeling

The key issue that arises in modeling the high-resolution image data generated in AYIN’s applications, is how to include large-scale spatial, temporal, and spectral dependencies. AYIN investigates approaches to the construction of image models including such dependencies. A central question in the use of such models is how to deal with the large data volumes arising both from the large size of the images involved, and the existence of large image collections. Fortunately, high dimensionality typically implies data redundancy, and so AYIN investigates methods for reducing the dimensionality of the data and describing the spatial, temporal, and spectral dependencies in ways that allow efficient data processing.

3.2.1. Markov random fields with long-range and higher-order interactions

One way to achieve large-scale dependencies is via explicit long-range interactions. MRFs with long-range interactions are also used in AYIN to model geometric spatial and temporal structure, and the techniques and algorithms developed there will also be applied to image modeling. In modeling image structures, however, other important properties, such as control of the relative phase of Fourier components, and spontaneous symmetry breaking, may also be required. These properties can only be achieved by higher-order interactions. These require specific techniques and algorithms, which are developed in parallel with the models.

3.2.2. Hierarchical models

Another way to achieve long-range dependencies is via shorter range interactions in a hierarchical structure. AYIN works on the development of models defined as a set of hierarchical image partitions represented by a binary forest structure. Key sub-problems include the development of multi-feature models of image regions as an ensemble of spectral, texture, geometrical, and classification features, where we search to optimize the ratio between discrimination capacity of the feature space and dimensionality of this space; and the development of similarity criteria between image regions, which would compute distances between regions in the designed feature space and would be data-driven and scale-independent. One way to proceed in the latter case consists in developing a composite kernel method, which would seek to project multi-feature data into a new space, where regions from different thematic categories become linearly or almost linearly separable. This involves developing kernel functions as a combination of basis kernels, and estimating kernel-based support vector machine parameters.

3.3. Algorithms

Computational techniques are necessary in order to extract the information of interest from the models. In addition, most models contain ‘nuisance parameters’, including the structure of the models themselves, that must be dealt with in some way. AYIN is interested in adapting and developing methods for solving these problems in cases where existing methods are inadequate.

3.3.1. Nuisance parameters and parameter estimation

In order to render the models operational, it is crucial to find some way to deal with nuisance parameters. In a Bayesian framework, the parameters must be integrated or marginalized out. Unfortunately, this is usually very difficult. Fortunately, Laplace’s method often provides a good approximation, in many cases being equivalent to classical maximum likelihood parameter estimation. Even these problems are not easy to solve, however, when dealing with complex, structured models. This is particularly true when it is necessary to estimate simultaneously both the information of interest and the parameters. AYIN is developing a number of different methods for dealing with nuisance parameters, corresponding to the diversity of modeling approaches.

3.3.2. Information extraction

Extracting the information of interest from any model involves making estimates based on various criteria, for example MAP, MPM, or MMSE. Computing these estimates often requires the solution of hard optimization problems. The complexity of many of the models to be developed within AYIN means that off-the-shelf algorithms and current techniques are often not capable of solving these problems. AYIN develops a diversity of algorithmic approaches adapted to the particular models developed.

GRAPHDECO Project-Team

3. Research Program

3.1. Introduction

Our research program is oriented around two main axes: 1) Computer-Assisted Design with Heterogeneous Representations and 2) Graphics with Uncertainty and Heterogeneous Content. These two axes are governed by a set of common fundamental goals, share many common methodological tools and are deeply intertwined in the development of applications.

3.1.1. Computer-Assisted Design with Heterogeneous Representations

Designers use a variety of visual representations to explore and communicate about a concept. Figure 2 illustrates some typical representations, including sketches, hand-made prototypes, 3D models, 3D printed prototypes or instructions.

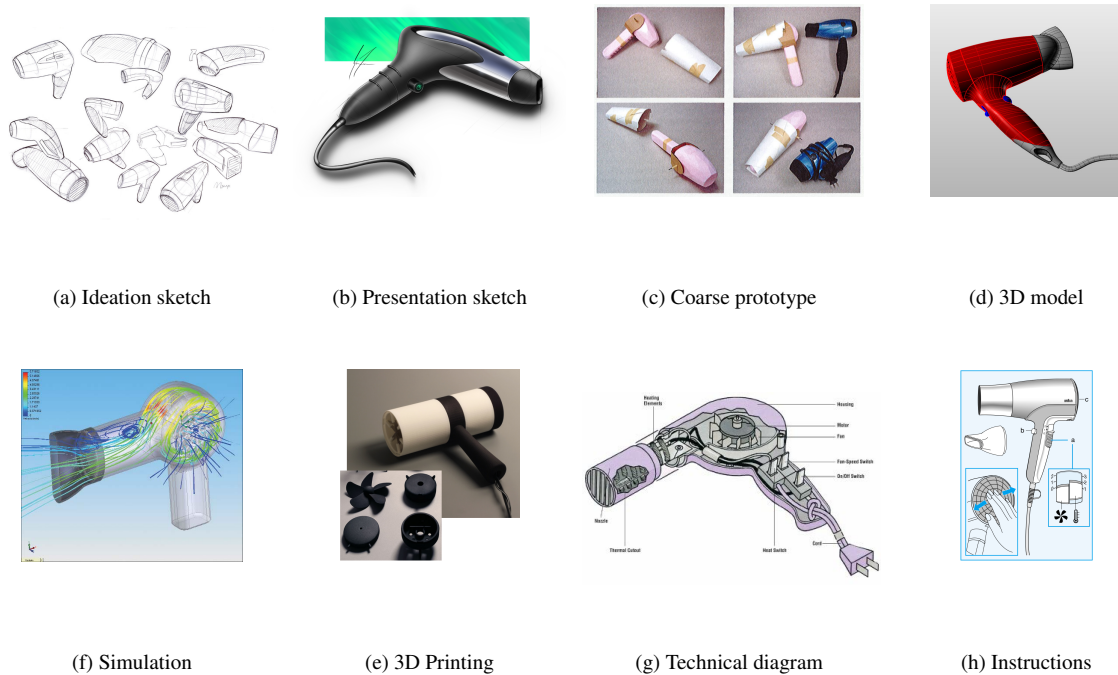


Figure 2. Various representations of a hair dryer at different stages of the design process. Image source, in order: *c-maeng on deviantart.com*, *shauntur on deviantart.com*, "Prototyping and Modelmaking for Product Design" Hallgrímsson, B., Laurence King Publishers, 2012, *samsher511 on turbosquid.com*, *my.solidworks.com*, *weilung tseng on cargocollective.com*, *howstuffworks.com*, *u-manual.com*

The early representations of a concept, such as rough sketches and hand-made prototypes, help designers formulate their ideas and test the form and function of multiple design alternatives. These low-fidelity representations are meant to be cheap and fast to produce, to allow quick exploration of the *design space* of the concept. These representations are also often approximate to leave room for subjective interpretation and to stimulate imagination; in this sense, these representations can be considered *uncertain*. As the concept gets more finalized, time and effort are invested in the production of more detailed and accurate representations, such as high-fidelity 3D models suitable for simulation and fabrication. These detailed models can also be used to create didactic instructions for assembly and usage.

Producing these different representations of a concept requires specific skills in sketching, modeling, manufacturing and visual communication. For these reasons, professional studios often employ different experts to produce the different representations of the same concept, at the cost of extensive discussions and numerous iterations between the actors of this process. The complexity of the multi-disciplinary skills involved in the design process also hinders their adoption by laymen.

Existing solutions to facilitate design have focused on a subset of the representations used by designers. However, no solution considers all representations at once, for instance to directly convert a series of sketches into a set of physical prototypes. In addition, all existing methods assume that the concept is unique rather than ambiguous. As a result, rich information about the variability of the concept is lost during each conversion step.

We plan to facilitate design for professionals and laymen by addressing the following objectives:

- We want to assist designers in the exploration of the *design space* that captures the possible variations of a concept. By considering a concept as a *distribution* of shapes and functionalities rather than a single object, our goal is to help designers consider multiple design alternatives more quickly and effectively. Such a representation should also allow designers to preserve multiple alternatives along all steps of the design process rather than committing to a single solution early on and pay the price of this decision for all subsequent steps. We expect that preserving alternatives will facilitate communication with engineers, managers and clients, accelerate design iterations and even allow mass personalization by the end consumers.
- We want to support the various representations used by designers during concept development. While drawings and 3D models have received significant attention in past Computer Graphics research, we will also account for the various forms of rough physical prototypes made to evaluate the shape and functionality of a concept. Depending on the task at hand, our algorithms will either analyse these prototypes to generate a virtual concept, or assist the creation of these prototypes from a virtual model. We also want to develop methods capable of adapting to the different drawing and manufacturing techniques used to create sketches and prototypes. We envision design tools that conform to the habits of users rather than impose specific techniques to them.
- We want to make professional design techniques available to novices. Affordable software, hardware and online instructions are democratizing technology and design, allowing small businesses and individuals to compete with large companies. New manufacturing processes and online interfaces also allow customers to participate in the design of an object via mass personalization. However, similarly to what happened for desktop publishing thirty years ago, desktop manufacturing tools need to be simplified to account for the needs and skills of novice designers. We hope to support this trend by adapting the techniques of professionals and by automating the tasks that require significant expertise.

3.1.2. Graphics with Uncertainty and Heterogeneous Content

Our research is motivated by the observation that traditional CG algorithms have not been designed to account for uncertain data. For example, global illumination rendering assumes accurate virtual models of geometry, light and materials to simulate light transport. While these algorithms produce images of high realism, capturing effects such as shadows, reflections and interreflections, they are not applicable to the growing mass of uncertain data available nowadays.

The need to handle uncertainty in CG is timely and pressing, given the large number of *heterogeneous sources of 3D content* that have become available in recent years. These include data from cheap depth+image sensors (e.g., Kinect or the Tango), 3D reconstructions from image/video data, but also data from large 3D geometry databases, or casual 3D models created using simplified sketch-based modeling tools. Such alternate content has varying levels of *uncertainty* about the scene or objects being modelled. This includes uncertainty in geometry, but also in materials and/or lights – which are often not even available with such content. Since CG algorithms cannot be applied directly, visual effects artists spend hundreds of hours correcting inaccuracies and completing the captured data to make them useable in film and advertising.

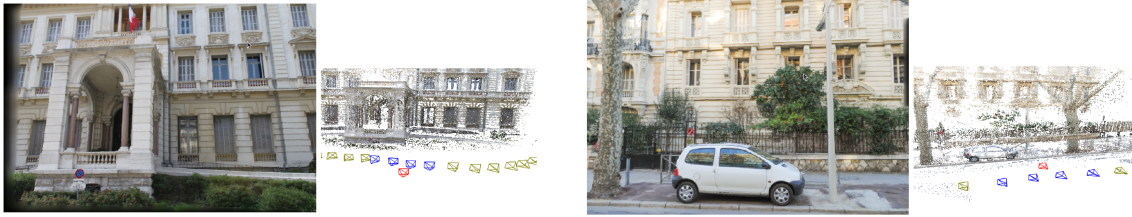


Figure 3. Image-Based Rendering (IBR) techniques use input photographs and approximate 3D to produce new synthetic views.

We identify a major scientific bottleneck which is the need to treat *heterogeneous* content, i.e., containing both (mostly captured) uncertain and perfect, traditional content. Our goal is to provide solutions to this bottleneck, by explicitly and formally modeling uncertainty in CG, and to develop new algorithms that are capable of mixed rendering for this content.

We strive to develop methods in which heterogeneous – and often uncertain – data can be handled automatically in CG with a principled methodology. Our main focus is on *rendering* in CG, including dynamic scenes (video/animations).

Given the above, we need to address the following challenges:

- Develop a theoretical model to handle uncertainty in computer graphics. We must define a new formalism that inherently incorporates uncertainty, and must be able to express traditional CG rendering, both physically accurate and approximate approaches. Most importantly, the new formulation must elegantly handle mixed rendering of perfect synthetic data and captured uncertain content. An important element of this goal is to incorporate *cost* in the choice of algorithm and the optimizations used to obtain results, e.g., preferring solutions which may be slightly less accurate, but cheaper in computation or memory.
- The development of rendering algorithms for heterogeneous content often requires preprocessing of image and video data, which sometimes also includes depth information. An example is the decomposition of images into intrinsic layers of reflectance and lighting, which is required to perform relighting. Such solutions are also useful as image-manipulation or computational photography techniques. The challenge will be to develop such “intermediate” algorithms for the uncertain and heterogeneous data we target.
- Develop efficient rendering algorithms for uncertain and heterogeneous content, reformulating rendering in a probabilistic setting where appropriate. Such methods should allow us to develop approximate rendering algorithms using our formulation in a well-grounded manner. The formalism should include probabilistic models of how the scene, the image and the data interact. These models should be data-driven, e.g., building on the abundance of online geometry and image databases, domain-driven, e.g., based on requirements of the rendering algorithms or perceptually guided, leading to plausible solutions based on limitations of perception.

GRAPHIK Project-Team

3. Research Program

3.1. Logic-based Knowledge Representation and Reasoning

We follow the mainstream *logic-based* approach to the KR domain. First-order logic (FOL) is the reference logic in KR and most formalisms in this area can be translated into fragments (i.e., particular subsets) of FOL. This is in particular the case for description logics and existential rules, two well-known KR formalisms studied in the team.

A large part of research in this domain can be seen as studying the *trade-off* between the expressivity of languages and the complexity of (sound and complete) reasoning in these languages. The fundamental problem in KR languages is entailment checking: is a given piece of knowledge entailed by other pieces of knowledge, for instance from a knowledge base (KB)? Another important problem is *consistency* checking: is a set of knowledge pieces (for instance the knowledge base itself) consistent, i.e., is it sure that nothing absurd can be entailed from it? The *ontology-mediated query answering* problem is a topical problem (see Section 3.3). It asks for the set of answers to a query in the KB. In the case of Boolean queries (i.e., queries with a yes/no answer), it can be recast as entailment checking.

3.2. Graph-based Knowledge Representation and Reasoning

Besides logical foundations, we are interested in KR formalisms that comply, or aim at complying with the following requirements: to have good *computational* properties and to allow users of knowledge-based systems to have a maximal *understanding and control* over each step of the knowledge base building process and use.

These two requirements are the core motivations for our graph-based approach to KR. We view labelled graphs as an *abstract representation* of knowledge that can be expressed in many KR languages (different kinds of conceptual graphs —historically our main focus—, the Semantic Web language RDF (Resource Description Framework), its extension RDFS (RDF Schema), expressive rules equivalent to the so-called tuple-generating-dependencies in databases, some description logics dedicated to query answering, etc.). For these languages, reasoning can be based on the structure of objects, thus based on graph-theoretic notions, while staying logically founded.

More precisely, our basic objects are labelled graphs (or hypergraphs) representing entities and relationships between these entities. These graphs have a natural translation in first-order logic. Our basic reasoning tool is graph homomorphism. The fundamental property is that graph homomorphism is sound and complete with respect to logical entailment *i.e.*, given two (labelled) graphs G and H , there is a homomorphism from G to H if and only if the formula assigned to G is entailed by the formula assigned to H . In other words, logical reasoning on these graphs can be performed by graph mechanisms. These knowledge constructs and the associated reasoning mechanisms can be extended (to represent rules for instance) while keeping this fundamental correspondence between graphs and logics.

3.3. Ontology-Mediated Query Answering

Querying knowledge bases has become a central problem in knowledge representation and in databases. A knowledge base (KB) is classically composed of a terminological part (metadata, ontology) and an assertional part (facts, data). Queries are supposed to be at least as expressive as the basic queries in databases, i.e., conjunctive queries, which can be seen as existentially closed conjunctions of atoms or as labelled graphs. The challenge is to define good trade-offs between the expressivity of the ontological language and the complexity of querying data in presence of ontological knowledge. Description logics have been so far the prominent family of formalisms for representing and reasoning with ontological knowledge. However, classical description logics were not designed for efficient data querying. On the other hand, database languages are able to process complex queries on huge databases, but without taking the ontology into account. There is thus a need for new languages and mechanisms, able to cope with the ever growing size of knowledge bases in the Semantic Web or in scientific domains.

This problem is related to two other problems identified as fundamental in KR:

- *Query-answering with incomplete information.* Incomplete information means that it might be unknown whether a given assertion is true or false. Databases classically make the so-called closed-world assumption: every fact that cannot be retrieved or inferred from the base is assumed to be false. Knowledge bases classically make the open-world assumption: if something cannot be inferred from the base, and neither can its negation, then its truth status is unknown. The need of coping with incomplete information is a distinctive feature of querying knowledge bases with respect to querying classical databases (however, as explained above, this distinction tends to disappear). The presence of incomplete information makes the query answering task much more difficult.
- *Reasoning with rules.* Researching types of rules and adequate manners to process them is a mainstream topic in the Semantic Web, and, more generally a crucial issue for knowledge-based systems. For several years, we have been studying some rules, both in their logical and their graph form, which are syntactically very simple but also very expressive. These rules, known as existential rules or Datalog+, can be seen as an abstraction of ontological knowledge expressed in the main languages used in the context of KB querying. See Section 6.1 for details on the results obtained.

A problem generalizing the above described problems, and particularly relevant in the context of multiple data/metadata sources, is *querying hybrid knowledge bases*. In a hybrid knowledge base, each component may have its own formalism and its own reasoning mechanisms. There may be a common ontology shared by all components, or each component may have its own ontology, with mappings being defined among the ontologies. The question is what kind of interactions between these components and/or what limitations on the languages preserve the decidability of basic problems and if so, a “reasonable” complexity. Note that there are strong connections with the issue of data integration in databases.

3.4. Imperfect Information and Priorities

While classical FOL is the kernel of many KR languages, to solve real-world problems we often need to consider features that cannot be expressed purely (or not naturally) in classical logic. The logic- and graph-based formalisms used for previous points have thus to be extended with such features. The following requirements have been identified from scenarios in decision making in the agronomy domain.

1. to cope with vague and uncertain information and preferences in queries;
2. to cope with multi-granularity knowledge;
3. to take into account different and potentially conflicting viewpoints ;
4. to integrate decision notions (priorities, gravity, risk, benefit);
5. to integrate argumentation-based reasoning.

Although the solutions we develop need to be validated on the applications that motivated them, we also want them to be sufficiently generic to be applied in other contexts. One angle of attack (but not the only possible one) consists in increasing the expressivity of our core languages, while trying to preserve their essential combinatorial properties, so that algorithmic optimizations can be transferred to these extensions.

HEPHAISTOS Project-Team

3. Research Program

3.1. Interval analysis

We are interested in real-valued system solving ($f(X) = 0$, $f(X) \leq 0$), in optimization problems, and in the proof of the existence of properties (for example, it exists X such that $f(X) = 0$ or it exist two values X_1, X_2 such that $f(X_1) > 0$ and $f(X_2) < 0$). There are few restrictions on the function f as we are able to manage explicit functions using classical mathematical operators (e.g. $\sin(x + y) + \log(\cos(e^x) + y^2)$) as well as implicit functions (e.g. determining if there are parameter values of a parametrized matrix such that the determinant of the matrix is negative, without calculating the analytical form of the determinant).

Solutions are searched within a finite domain (called a *box*) which may be either continuous or mixed (i.e. for which some variables must belong to a continuous range while other variables may only have values within a discrete set). An important point is that we aim at finding all the solutions within the domain whenever the computer arithmetic will allow it: in other words we are looking for *certified* solutions. For example, for 0-dimensional system solving, we will provide a box that contains one, and only one, solution together with a numerical approximation of this solution. This solution may further be refined at will using multi-precision.

The core of our methods is the use of *interval analysis* that allows one to manipulate mathematical expressions whose unknowns have interval values. A basic component of interval analysis is the *interval evaluation* of an expression. Given an analytical expression F in the unknowns $\{x_1, x_2, \dots, x_n\}$ and ranges $\{X_1, X_2, \dots, X_n\}$ for these unknowns we are able to compute a range $[A, B]$, called the interval evaluation, such that

$$\forall \{x_1, x_2, \dots, x_n\} \in \{X_1, X_2, \dots, X_n\}, A \leq F(x_1, x_2, \dots, x_n) \leq B \quad (6)$$

In other words the interval evaluation provides a lower bound of the minimum of F and an upper bound of its maximum over the box.

For example if $F = x \sin(x + x^2)$ and $x \in [0.5, 1.6]$, then $F([0.5, 1.6]) = [-1.362037441, 1.6]$, meaning that for any x in $[0.5, 1.6]$ we guarantee that $-1.362037441 \leq f(x) \leq 1.6$.

The interval evaluation of an expression has interesting properties:

- it can be implemented in such a way that the results are guaranteed with respect to round-off errors i.e. property 1 is still valid in spite of numerical errors induced by the use of floating point numbers
- if $A > 0$ or $B < 0$, then no values of the unknowns in their respective ranges can cancel F
- if $A > 0$ ($B < 0$), then F is positive (negative) for any value of the unknowns in their respective ranges

A major drawback of the interval evaluation is that $A(B)$ may be overestimated i.e. values of x_1, x_2, \dots, x_n such that $F(x_1, x_2, \dots, x_n) = A(B)$ may not exist. This overestimation occurs because in our calculation each occurrence of a variable is considered as an independent variable. Hence if a variable has multiple occurrences, then an overestimation may occur. Such phenomena can be observed in the previous example where $B = 1.6$ while the real maximum of F is approximately 0.9144. The value of B is obtained because we are using in our calculation the formula $F = x \sin(y + z^2)$ with y, z having the same interval value than x .

Fortunately there are methods that allow one to reduce the overestimation and the overestimation amount decreases with the width of the ranges. The latter remark leads to the use of a branch-and-bound strategy in which for a given box a variable range will be bisected, thereby creating two new boxes that are stored in a list and processed later on. The algorithm is complete if all boxes in the list have been processed, or if during the process a box generates an answer to the problem at hand (e.g. if we want to prove that $F(X) < 0$, then the algorithm stops as soon as $F(\mathcal{B}) \geq 0$ for a certain box \mathcal{B}).

A generic interval analysis algorithm involves the following steps on the current box [1], [8], [5]:

1. *exclusion operators*: these operators determine that there is no solution to the problem within a given box. An important issue here is the extensive and smart use of the monotonicity of the functions
2. *filters*: these operators may reduce the size of the box i.e. decrease the width of the allowed ranges for the variables
3. *existence operators*: they allow one to determine the existence of a unique solution within a given box and are usually associated with a numerical scheme that allows for the computation of this solution in a safe way
4. *bisection*: choose one of the variable and bisect its range for creating two new boxes
5. *storage*: store the new boxes in the list

The scope of the HEPHAISTOS project is to address all these steps in order to find the most efficient procedures. Our efforts focus on mathematical developments (adapting classical theorems to interval analysis, proving interval analysis theorems), the use of symbolic computation and formal proofs (a symbolic pre-processing allows one to automatically adapt the solver to the structure of the problem), software implementation and experimental tests (for validation purposes).

Important note: We have insisted on interval analysis because this is a **major component** of our robotics activity. Our theoretical work in robotics is an analysis of the robotic environment in order to exhibit proofs on the behavior of the system that may be qualitative (e.g. the proof that a cable-driven parallel robot with more than 6 non-deformable cables will have at most 6 cables under tension simultaneously) or quantitative. In the quantitative case as we are dealing with realistic and not toy examples (including our own prototypes that are developed whenever no equivalent hardware is available or to verify our assumptions) we have to manage problems that are so complex that analytical solutions are probably out of reach (e.g. the direct kinematics of parallel robots) and we have to resort to algorithms and numerical analysis. We are aware of different approaches in numerical analysis (e.g. some team members were previously involved in teams devoted to computational geometry and algebraic geometry) but interval analysis provides us another approach with high flexibility, the possibility of managing non algebraic problems (e.g. the kinematics of cable-driven parallel robots with sagging cables, that involves inverse hyperbolic functions) and to address various types of issues (system solving, optimization, proof of existence ...).

3.2. Robotics

HEPHAISTOS, as a follow-up of COPRIN, has a long-standing tradition of robotics studies, especially for closed-loop robots [4], especially cable-driven parallel robots. We address theoretical issues with the purpose of obtaining analytical and theoretical solutions, but in many cases only numerical solutions can be obtained due to the complexity of the problem. This approach has motivated the use of interval analysis for two reasons:

1. the versatility of interval analysis allows us to address issues (e.g. singularity analysis) that cannot be tackled by any other method due to the size of the problem
2. uncertainties (which are inherent to a robotic device) have to be taken into account so that the *real* robot is guaranteed to have the same properties as the *theoretical* one, even in the worst case [15]. This is a crucial issue for many applications in robotics (e.g. medical or assistance robot)

Our field of study in robotics focuses on *kinematic* issues such as workspace and singularity analysis, positioning accuracy, trajectory planning, reliability, calibration, modularity management and, prominently, *appropriate design*, i.e. determining the dimensioning of a robot mechanical architecture that guarantees that the real robot satisfies a given set of requirements. The methods that we develop can be used for other robotic problems, see for example the management of uncertainties in aircraft design [6].

Our theoretical work must be validated through experiments that are essential for the sake of credibility. A contrario, experiments will feed theoretical work. Hence HEPHAISTOS works with partners on the development of real robots but also develops its own prototypes. In the last years we have developed a large number of prototypes and we have extended our development to devices that are not strictly robots but are part of an overall environment for assistance. We benefit here from the development of new miniature, low energy computers with an interface for analog and logical sensors such as the Arduino or the Phidgets. The web pages <http://www-sop.inria.fr/hephaistos/mediatheque/index.html> presents all of our prototypes and experimental work.

LAGADIC Project-Team

3. Research Program

3.1. Visual servoing

Basically, visual servoing techniques consist in using the data provided by one or several cameras in order to control the motions of a dynamic system [1]. Such systems are usually robot arms, or mobile robots, but can also be virtual robots, or even a virtual camera. A large variety of positioning tasks, or mobile target tracking, can be implemented by controlling from one to all the degrees of freedom of the system. Whatever the sensor configuration, which can vary from one on-board camera on the robot end-effector to several free-standing cameras, a set of visual features has to be selected at best from the image measurements available, allowing to control the desired degrees of freedom. A control law has also to be designed so that these visual features $\mathbf{s}(t)$ reach a desired value \mathbf{s}^* , defining a correct realization of the task. A desired planned trajectory $\mathbf{s}^*(t)$ can also be tracked. The control principle is thus to regulate the error vector $\mathbf{s}(t) - \mathbf{s}^*(t)$ to zero. With a vision sensor providing 2D measurements, potential visual features are numerous, since 2D data (coordinates of feature points in the image, moments, ...) as well as 3D data provided by a localization algorithm exploiting the extracted 2D features can be considered. It is also possible to combine 2D and 3D visual features to take the advantages of each approach while avoiding their respective drawbacks.

More precisely, a set \mathbf{s} of k visual features can be taken into account in a visual servoing scheme if it can be written:

$$\mathbf{s} = \mathbf{s}(\mathbf{x}(\mathbf{p}(t)), \mathbf{a}) \quad (7)$$

where $\mathbf{p}(t)$ describes the pose at the instant t between the camera frame and the target frame, \mathbf{x} the image measurements, and \mathbf{a} a set of parameters encoding a potential additional knowledge, if available (such as for instance a coarse approximation of the camera calibration parameters, or the 3D model of the target in some cases).

The time variation of \mathbf{s} can be linked to the relative instantaneous velocity \mathbf{v} between the camera and the scene:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{p}} \dot{\mathbf{p}} = \mathbf{L}_s \mathbf{v} \quad (8)$$

where \mathbf{L}_s is the interaction matrix related to \mathbf{s} . This interaction matrix plays an essential role. Indeed, if we consider for instance an eye-in-hand system and the camera velocity as input of the robot controller, we obtain when the control law is designed to try to obtain an exponential decoupled decrease of the error:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L}}_s^+ (\mathbf{s} - \mathbf{s}^*) - \widehat{\mathbf{L}}_s^+ \frac{\partial \mathbf{s}}{\partial t} \quad (9)$$

where λ is a proportional gain that has to be tuned to minimize the time-to-convergence, $\widehat{\mathbf{L}}_s^+$ is the pseudo-inverse of a model or an approximation of the interaction matrix, and $\frac{\partial \mathbf{s}}{\partial t}$ an estimation of the features velocity due to a possible own object motion.

From the selected visual features and the corresponding interaction matrix, the behavior of the system will have particular properties as for stability, robustness with respect to noise or to calibration errors, robot 3D trajectory, etc. Usually, the interaction matrix is composed of highly non linear terms and does not present any decoupling properties. This is generally the case when s is directly chosen as x . In some cases, it may lead to inadequate robot trajectories or even motions impossible to realize, local minimum, tasks singularities, etc. It is thus extremely important to design adequate visual features for each robot task or application, the ideal case (very difficult to obtain) being when the corresponding interaction matrix is constant, leading to a simple linear control system. To conclude in a few words, **visual servoing is basically a non linear control problem. Our Holy Grail quest is to transform it into a linear control problem.**

Furthermore, embedding visual servoing in the task function approach allows solving efficiently the redundancy problems that appear when the visual task does not constrain all the degrees of freedom of the system. It is then possible to realize simultaneously the visual task and secondary tasks such as visual inspection, or joint limits or singularities avoidance. This formalism can also be used for tasks sequencing purposes in order to deal with high level complex applications.

3.2. Visual tracking

Elaboration of object tracking algorithms in image sequences is an important issue for researches and applications related to visual servoing and more generally for robot vision. A robust extraction and real time spatio-temporal tracking process of visual cues is indeed one of the keys to success of a visual servoing task. If fiducial markers may still be useful to validate theoretical aspects in modeling and control, natural scenes with non-cooperative objects and subject to various illumination conditions have to be considered for addressing large scale realistic applications.

Most of the available tracking methods can be divided into two main classes: feature-based and model-based. The former approach focuses on tracking 2D features such as geometrical primitives (points, segments, circles,...), object contours, regions of interest, etc. The latter explicitly uses a model of the tracked objects. This can be either a 3D model or a 2D template of the object. This second class of methods usually provides a more robust solution. Indeed, the main advantage of the model-based methods is that the knowledge about the scene allows improving tracking robustness and performance, by being able to predict hidden movements of the object, detect partial occlusions and acts to reduce the effects of outliers. The challenge is to build algorithms that are fast and robust enough to meet our application requirements. Therefore, even if we still consider 2D feature tracking in some cases, our researches mainly focus on real-time 3D model-based tracking, since these approaches are very accurate, robust, and well adapted to any class of visual servoing schemes. Furthermore, they also meet the requirements of other classes of application, such as augmented reality.

3.3. Slam

Most of the applications involving mobile robotic systems (ground vehicles, aerial robots, automated submarines,...) require a reliable localization of the robot in its environment. A challenging problem is when neither the robot localization nor the map is known. Localization and mapping must then be considered concurrently. This problem is known as Simultaneous Localization And Mapping (Slam). In this case, the robot moves from an unknown location in an unknown environment and proceeds to incrementally build up a navigation map of the environment, while simultaneously using this map to update its estimated position.

Nevertheless, solving the Slam problem is not sufficient for guaranteeing an autonomous and safe navigation. The choice of the representation of the map is, of course, essential. The representation has to support the different levels of the navigation process: motion planning, motion execution and collision avoidance and, at the global level, the definition of an optimal strategy of displacement. The original formulation of the Slam problem is purely metric (since it basically consists in estimating the Cartesian situations of the robot and a set of landmarks), and it does not involve complex representations of the environment. However, it is now well recognized that **several complementary representations are needed to perform exploration, navigation, mapping, and control tasks successfully. We propose to use composite models of the environment that**

mix topological, metric, and grid-based representations. Each type of representation is well adapted to a particular aspect of autonomous navigation [7]: the metric model allows one to locate the robot precisely and plan Cartesian paths, the topological model captures the accessibility of different sites in the environment and allows a coarse localization, and finally the grid representation is useful to characterize the free space and design potential functions used for reactive obstacle avoidance. However, ensuring the consistency of these various representations during the robot exploration, and merging observations acquired from different viewpoints by several cooperative robots, are difficult problems. This is particularly true when different sensing modalities are involved. New studies to derive efficient algorithms for manipulating the hybrid representations (merging, updating, filtering...) while preserving their consistency are needed.

3.4. Scene modeling and understanding

Long-term mapping has received an increasing amount of attention during last years, largely motivated by the growing need to integrate robots into the real world wherein dynamic objects constantly change the appearance of the scene. A mobile robot evolving in such a dynamic world should not only be able to build a map of the observed environment at a specific moment, but also to maintain this map consistent over a long period of time. It has to deal with dynamic changes that can cause the navigation process to fail. However updating the map is particularly challenging in large-scale environments. To identify changes, robots have to keep a memory of the previous states of the environment and the more dynamic it is, the higher will be the number of states to manage and the more computationally intensive will be the updating process. Mapping large-scale dynamic environments is then particularly difficult as the map size can be arbitrary large. Additionally, mapping many times the whole environment is not always possible or convenient and it is useful to take advantages of methods using only a small number of observations.

A recent trend in robotic mapping is to augment low-level maps with semantic interpretation of their content, which allows to improve the robot's environmental awareness through the use of high-level concepts. In mobile robot navigation, the so-called semantic maps have already been used to improve path planning methods, mainly by providing the robot with the ability to deal with human-understandable targets.

STARS Project-Team

3. Research Program

3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved:** *the software engineering* research direction provides new methodologies for building safe activity recognition systems and *the perception* and *the semantic activity recognition* directions provide new activity recognition techniques which are designed and validated for concrete video analytic and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversely, we consider a *new research axis in machine learning*, combining a priori knowledge and learning techniques, to set up the various models of an activity recognition system. A major objective is to automate model building or model enrichment at the perception level and at the understanding level.

3.2. Perception for Activity Recognition

Participants: François Brémond, Sabine Moisan, Monique Thonnat.

Computer Vision; Cognitive Systems; Learning; Activity Recognition.

3.2.1. Introduction

Our main goal in perception is to develop vision algorithms able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives. We have also several issues related to perception which combine machine learning and perception techniques: learning people appearance, parameters for system control and shape statistics.

3.2.2. Appearance Models and People Tracking

An important issue is to detect in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios.

Appearance models. In particular, we study the temporal variation of the features characterizing the appearance of a human. This task could be achieved by clustering potential candidates depending on their position and their reliability. This task can provide any people tracking algorithms with reliable features allowing for instance to (1) better track people or their body parts during occlusion, or to (2) model people appearance for re-identification purposes in mono and multi-camera networks, which is still an open issue. The underlying challenge of the person re-identification problem arises from significant differences in illumination, pose and camera parameters. The re-identification approaches have two aspects: (1) establishing correspondences between body parts and (2) generating signatures that are invariant to different color responses. As we have already several descriptors which are color invariant, we now focus more on aligning two people detection and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. As the results of the re-identification are always given as the ranking list, re-identification focuses on learning to rank. "Learning to rank" is a type of machine learning problem, in which the goal is to automatically construct a ranking model from a training data.

Therefore, we work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). New 3D RGB-D sensors are also investigated, to help in getting an accurate segmentation for specific scene conditions.

Long term tracking. For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modeling and managing the uncertainty of these processes is also an open issue. In Stars we propose to add a reasoning layer to a classical Bayesian framework modeling the uncertainty of the tracked objects. This reasoning layer can take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

Controlling system parameters. Another research direction is to manage a library of video processing programs. We are building a perception library by selecting robust algorithms for feature extraction, by insuring they work efficiently with real time constraints and by formalizing their conditions of use within a program supervision model. In the case of video cameras, at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues, we are developing new learning techniques.

3.3. Semantic Activity Recognition

Participants: François Brémond, Sabine Moisan, Monique Thonnat.

Activity Recognition, Scene Understanding, Computer Vision

3.3.1. Introduction

Semantic activity recognition is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and activities. The signal and the feature levels are characterized by strong noise, ambiguous, corrupted and missing data. The whole process of scene understanding consists in analyzing this information to bring forth pertinent insight of the scene and its dynamics while handling the low level noise. Moreover, to obtain a semantic abstraction, building activity models is a crucial point. A still open issue consists in determining whether these models should be given a priori or learned. Another challenge consists in organizing this knowledge in order to capitalize experience, share it with others and update it along with experimentation. To face this challenge, tools in knowledge engineering such as machine learning or ontology are needed.

Thus we work along the following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models), learning (how to learn the models needed for activity recognition) and activity recognition and discrete event systems.

3.3.2. High Level Understanding

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event).

Concerning the modeling of activities, we are working towards two directions: the uncertainty management for representing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. For the second direction, we built a language for video event modeling and a visual concept ontology (including color, texture and spatial concepts) to be extended with temporal concepts (motion, trajectories, events ...) and other perceptual concepts (physiological sensor concepts ...).

3.3.3. Learning for Activity Recognition

Given the difficulty of building an activity recognition system with a priori knowledge for a new application, we study how machine learning techniques can automate building or completing models at the perception level and at the understanding level.

At the understanding level, we are learning primitive event detectors. This can be done for example by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual feature samples. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. leveraging the human annotation task). A second direction is to learn typical composite event models for frequent activities using trajectory clustering or data mining techniques. We name composite event a particular combination of several primitive events.

3.3.4. Activity Recognition and Discrete Event Systems

The previous research axes are unavoidable to cope with the semantic interpretations. However they tend to let aside the pure event driven aspects of scenario recognition. These aspects have been studied for a long time at a theoretical level and led to methods and tools that may bring extra value to activity recognition, the most important being the possibility of formal analysis, verification and validation.

We have thus started to specify a formal model to define, analyze, simulate, and prove scenarios. This model deals with both absolute time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension, and verification). Our purpose is to offer a generic tool to express and recognize activities associated with a concrete language to specify activities in the form of a set of scenarios with temporal constraints. The theoretical foundations and the tools being shared with Software Engineering aspects, they will be detailed in section 3.4 .

The results of the research performed in perception and semantic activity recognition (first and second research directions) produce new techniques for scene understanding and contribute to specify the needs for new software architectures (third research direction).

3.4. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, François Brémont.

Software Engineering, Generic Components, Knowledge-based Systems, Software Component Platform, Object-oriented Frameworks, Software Reuse, Model-driven Engineering

The aim of this research axis is to build general solutions and tools to develop systems dedicated to activity recognition. For this, we rely on state-of-the art Software Engineering practices to ensure both sound design and easy use, providing genericity, modularity, adaptability, reusability, extensibility, dependability, and maintainability.

This research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. We work on the following three research axes: *models* (adapted to the activity recognition domain), *platform architecture* (to cope with deployment constraints and run time adaptation), and *system verification* (to generate dependable systems). For all these tasks we follow state of the art Software Engineering practices and, if needed, we attempt to set up new ones.

3.4.1. Platform Architecture for Activity Recognition

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [14], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system.

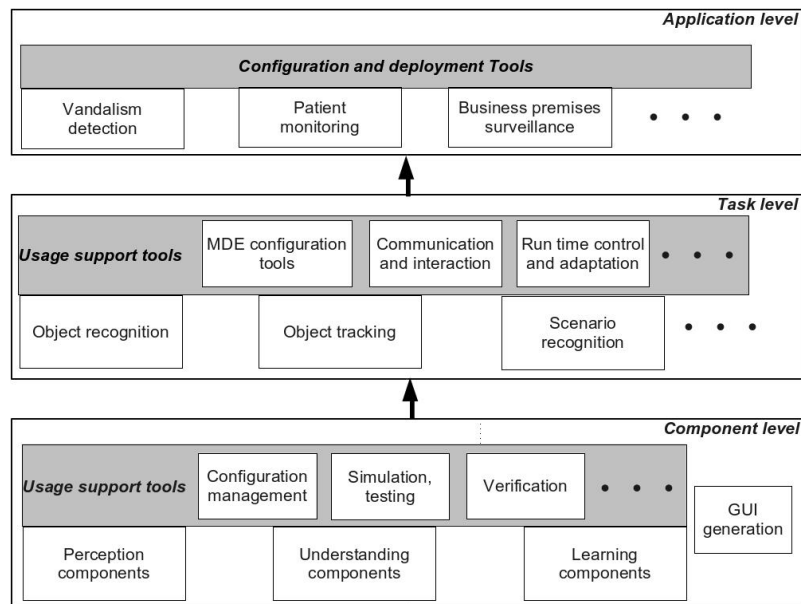


Figure 4. Global Architecture of an Activity Recognition The gray areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

Figure 4 presents our conceptual vision for the architecture of an activity recognition platform. It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.
 - *Perception components* contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
 - *Understanding components* provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.
 - *Learning components* implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

An Activity Recognition system is likely to pick components from these three packages. Hence, tools must be provided to configure (select, assemble), simulate, verify the resulting component combination. Other support tools may help to generate task or application dedicated languages or graphic interfaces.

- The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate in a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Object Recognition, Tracking, Scenario Recognition... In the future, other tasks will probably enrich this level.

For these tasks to nicely collaborate, communication and interaction facilities are needed. We shall also add MDE-enhanced tools for configuration and run-time adaptation.

- The **Application Level** integrates several of these tasks to build a system for a particular type of application, e.g., vandalism detection, patient monitoring, aircraft loading/unloading surveillance, etc.. Each system is parameterized to adapt to its local environment (number, type, location of sensors, scene geometry, visual parameters, number of objects of interest...). Thus configuration and deployment facilities are required.

The philosophy of this architecture is to offer at each level a balance between the widest possible genericity and the maximum effective reusability, in particular at the code level.

To cope with real application requirements, we shall also investigate distributed architecture, real time implementation, and user interfaces.

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving for verification, etc. Note that, in Figure 4, some of the boxes can be naturally adapted from SUP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

3.4.2. *Discrete Event Models of Activities*

As mentioned in the previous section (3.3) we have started to specify a formal model of scenario dealing with both absolute time and logical time. Our scenario and time models as well as the platform verification tools rely on a formal basis, namely the synchronous paradigm. To recognize scenarios, we consider activity descriptions as synchronous reactive systems and we apply general modeling methods to express scenario behavior.

Activity recognition systems usually exhibit many safeness issues. From the software engineering point of view we only consider software security. Our previous work on verification and validation has to be pursued; in particular, we need to test its scalability and to develop associated tools. Model-checking is an appealing technique since it can be automatized and helps to produce a code that has been formally proved. Our verification method follows a compositional approach, a well-known way to cope with scalability problems in model-checking.

Moreover, recognizing real scenarios is not a purely deterministic process. Sensor performance, precision of image analysis, scenario descriptions may induce various kinds of uncertainty. While taking into account this uncertainty, we should still keep our model of time deterministic, modular, and formally verifiable. To formally describe probabilistic timed systems, the most popular approach involves probabilistic extension of timed automata. New model checking techniques can be used as verification means, but relying on model checking techniques is not sufficient. Model checking is a powerful tool to prove decidable properties but introducing uncertainty may lead to infinite state or even undecidable properties. Thus model checking validation has to be completed with non exhaustive methods such as abstract interpretation.

3.4.3. *Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems*

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [45]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify models to represent transformations. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [77]. In particular, the constraints can be analyzed as a SAT problem.

An additional challenge is to manage the possible run-time changes of implementation due to context variations (e.g., lighting conditions, changes in the reference scene, etc.). Video surveillance systems have to dynamically adapt to a changing environment. The use of models at run-time is a solution. We are defining adaptation rules corresponding to the dependency constraints between specification elements in one model and software variants in the other [44], [89], [82].

TITANE Project-Team

3. Research Program

3.1. Context

Geometric modeling and processing revolve around three main end goals: a computerized shape representation that can be visualized (creating a realistic or artistic depiction), simulated (anticipating the real) or realized (manufacturing a conceptual or engineering design). Aside from the mere editing of geometry, central research themes in geometric modeling involve conversions between physical (real), discrete (digital), and mathematical (abstract) representations. Going from physical to digital is referred to as shape acquisition and reconstruction; going from mathematical to discrete is referred to as shape approximation and mesh generation; going from discrete to physical is referred to as shape rationalization.

Geometric modeling has become an indispensable component for computational and reverse engineering. Simulations are now routinely performed on complex shapes issued not only from computer-aided design but also from an increasing amount of available measurements. The scale of acquired data is quickly growing: we no longer deal exclusively with individual shapes, but with entire *scenes*, possibly at the scale of entire cities, with many objects defined as structured shapes. We are witnessing a rapid evolution of the acquisition paradigms with an increasing variety of sensors and the development of community data, as well as disseminated data.

In recent years, the evolution of acquisition technologies and methods has translated in an increasing overlap of algorithms and data in the computer vision, image processing, and computer graphics communities. Beyond the rapid increase of resolution through technological advances of sensors and methods for mosaicing images, the line between laser scan data and photos is getting thinner. Combining, e.g., laser scanners with panoramic cameras leads to massive 3D point sets with color attributes. In addition, it is now possible to generate dense point sets not just from laser scanners but also from photogrammetry techniques when using a well-designed acquisition protocol. Depth cameras are getting increasingly common, and beyond retrieving depth information we can enrich the main acquisition systems with additional hardware to measure geometric information about the sensor and improve data registration: e.g., accelerometers or GPS for geographic location, and compasses or gyrometers for orientation. Finally, complex scenes can be observed at different scales ranging from satellite to pedestrian through aerial levels.

These evolutions allow practitioners to measure urban scenes at resolutions that were until now possible only at the scale of individual shapes. The related scientific challenge is however more than just dealing with massive data sets coming from increase of resolution, as complex scenes are composed of multiple objects with structural relationships. The latter relate i) to the way the individual shapes are grouped to form objects, object classes or hierarchies, ii) to geometry when dealing with similarity, regularity, parallelism or symmetry, and iii) to domain-specific semantic considerations. Beyond reconstruction and approximation, consolidation and synthesis of complex scenes require rich structural relationships.

The problems arising from these evolutions suggest that the strengths of geometry and images may be combined in the form of new methodological solutions such as photo-consistent reconstruction. In addition, the process of measuring the geometry of sensors (through gyrometers and accelerometers) often requires both geometry process and image analysis for improved accuracy and robustness. Modeling urban scenes from measurements illustrates this growing synergy, and it has become a central concern for a variety of applications ranging from urban planning to simulation through rendering and special effects.

3.2. Analysis

Complex scenes are usually composed of a large number of objects which may significantly differ in terms of complexity, diversity, and density. These objects must be identified and their structural relationships must be recovered in order to model the scenes with improved robustness, low complexity, variable levels of details and ultimately, semantization (automated process of increasing degree of semantic content).

Object classification is an ill-posed task in which the objects composing a scene are detected and recognized with respect to predefined classes, the objective going beyond scene segmentation. The high variability in each class may explain the success of the stochastic approach which is able to model widely variable classes. As it requires a priori knowledge this process is often domain-specific such as for urban scenes where we wish to distinguish between instances as ground, vegetation and buildings. Additional challenges arise when each class must be refined, such as roof super-structures for urban reconstruction.

Structure extraction consists in recovering structural relationships between objects or parts of object. The structure may be related to adjacencies between objects, hierarchical decomposition, singularities or canonical geometric relationships. It is crucial for effective geometric modeling through levels of details or hierarchical multiresolution modeling. Ideally we wish to learn the structural rules that govern the physical scene manufacturing. Understanding the main canonical geometric relationships between object parts involves detecting regular structures and equivalences under certain transformations such as parallelism, orthogonality and symmetry. Identifying structural and geometric repetitions or symmetries is relevant for dealing with missing data during data consolidation.

Data consolidation is a problem of growing interest for practitioners, with the increase of heterogeneous and defect-laden data. To be exploitable, such defect-laden data must be consolidated by improving the data sampling quality and by reinforcing the geometrical and structural relations sub-tending the observed scenes. Enforcing canonical geometric relationships such as local coplanarity or orthogonality is relevant for registration of heterogeneous or redundant data, as well as for improving the robustness of the reconstruction process.

3.3. Approximation

Our objective is to explore the approximation of complex shapes and scenes with surface and volume meshes, as well as on surface and domain tiling. A general way to state the shape approximation problem is to say that we search for the shape discretization (possibly with several levels of detail) that realizes the best complexity / distortion trade-off. Such a problem statement requires defining a discretization model, an error metric to measure distortion as well as a way to measure complexity. The latter is most commonly expressed in number of polygon primitives, but other measures closer to information theory lead to measurements such as number of bits or minimum description length.

For surface meshes we intend to conceive methods which provide control and guarantees both over the global approximation error and over the validity of the embedding. In addition, we seek for resilience to heterogeneous data, and robustness to noise and outliers. This would allow repairing and simplifying triangle soups with cracks, self-intersections and gaps. Another exploratory objective is to deal generically with different error metrics such as the symmetric Hausdorff distance, or a Sobolev norm which mixes errors in geometry and normals.

For surface and domain tiling the term meshing is substituted for tiling to stress the fact that tiles may be not just simple elements, but can model complex smooth shapes such as bilinear quadrangles. Quadrangle surface tiling is central for the so-called *resurfacing* problem in reverse engineering: the goal is to tile an input raw surface geometry such that the union of the tiles approximates the input well and such that each tile matches certain properties related to its shape or its size. In addition, we may require parameterization domains with a simple structure. Our goal is to devise surface tiling algorithms that are both reliable and resilient to defect-laden inputs, effective from the shape approximation point of view, and with flexible control upon the structure of the tiling.

3.4. Reconstruction

Assuming a geometric dataset made out of points or slices, the process of shape reconstruction amounts to recovering a surface or a solid that matches these samples. This problem is inherently ill-posed as infinitely-many shapes may fit the data. One must thus regularize the problem and add priors such as simplicity or smoothness of the inferred shape.

The concept of geometric simplicity has led to a number of interpolating techniques commonly based upon the Delaunay triangulation. The concept of smoothness has led to a number of approximating techniques that commonly compute an implicit function such that one of its isosurfaces approximates the inferred surface. Reconstruction algorithms can also use an explicit set of prior shapes for inference by assuming that the observed data can be described by these predefined prior shapes. One key lesson learned in the shape problem is that there is probably not a single solution which can solve all cases, each of them coming with its own distinctive features. In addition, some data sets such as point sets acquired on urban scenes are very domain-specific and require a dedicated line of research.

In recent years the *smooth, closed case* (i.e., shapes without sharp features nor boundaries) has received considerable attention. However, the state-of-the-art methods have several shortcomings: in addition to being in general not robust to outliers and not sufficiently robust to noise, they often require additional attributes as input, such as lines of sight or oriented normals. We wish to devise shape reconstruction methods which are both geometrically and topologically accurate without requiring additional attributes, while exhibiting resilience to defect-laden inputs. Resilience formally translates into stability with respect to noise and outliers. Correctness of the reconstruction translates into convergence in geometry and (stable parts of) topology of the reconstruction with respect to the inferred shape known through measurements.

Moving from the smooth, closed case to the *piecewise smooth case* (possibly with boundaries) is considerably harder as the ill-posedness of the problem applies to each sub-feature of the inferred shape. Further, very few approaches tackle the combined issue of robustness (to sampling defects, noise and outliers) and feature reconstruction.

WIMMICS Project-Team

3. Research Program

3.1. Users Modeling and Designing Interaction on the Web

Wimmics focuses on interactions of ordinary users with ontology-based knowledge systems with a preference for semantic Web formalisms and Web 2.0 applications. We specialize interaction design and evaluation methods to Web application tasks such as searching, browsing, contributing or protecting data. The team is especially interested in using semantics in assisting the interactions. We propose knowledge graph representations and algorithms to support interaction adaptation for instance for context-awareness or intelligent interactions with machine. We propose and evaluate Web-based visualization techniques for linked data, querying, reasoning, explaining and justifying. Wimmics also integrates natural language processing approaches to support natural language based interactions. We rely on cognitive studies to build models of the system, the user and the interactions between users through the system, in order to support and improve these interactions. We extend the user modeling technique known as *Personas* where user models are represented as specific, individual humans. *Personas* are derived from significant behavior patterns (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. Our user models specialize *Personas* approaches to include aspects appropriate to Web applications. Wimmics also extends user models to capture very different aspects (e.g. emotional states).

3.2. Communities and Social Interactions Analysis

The domain of social network analysis is a whole research domain in itself and Wimmics targets what can be done with typed graphs, knowledge representations and social models. We also focus on the specificity of social Web and semantic Web applications and in bridging and combining the different social Web data structures and semantic Web formalisms. Beyond the individual user models, we rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where and when formal semantics is useful. We propose models of collectives of users and of their collaborative functioning extending the collaboration personas and methods to assess the quality of coordination interactions and the quality of coordination artifacts. We extend and compare community detection algorithms to identify and label communities of interest with the topics they share. We propose mixed representations containing social semantic representations (e.g. folksonomies) and formal semantic representations (e.g. ontologies) and propose operations that allow us to couple them and exchange knowledge between them. Moving to social interaction we develop models and algorithms to mine and integrate different yet linked aspects of social media contributions (opinions, arguments and emotions) relying in particular on natural language processing and argumentation theory. To complement the study of communities we rely on multi-agent systems to simulate and study social behaviors. Finally we also rely on Web 2.0 principles to provide and evaluate social Web applications.

3.3. Vocabularies, Semantic Web and Linked Data Based Knowledge Representation

For all the models we identified in the previous sections, we rely on and evaluate knowledge representation methodologies and theories, in particular ontology-based modeling. We also propose models and formalisms to capture and merge representations of different levels of semantics (e.g. formal ontologies and social folksonomies). The important point is to allow us to capture those structures precisely and flexibly and yet create as many links as possible between these different objects. We propose vocabularies and semantic Web formalizations for the whole aspects we model and we consider and study extensions of these formalisms when needed. The results have all in common to pursue the representation and publication of our models as linked

data. We also contribute to the transformation and linking of existing resources (informal models, databases, texts, etc.) to be published on the semantic Web and as linked data. Examples of aspects we formalize include: user profiles, social relations, linguistic knowledge, business processes, derivation rules, temporal descriptions, explanations, presentation conditions, access rights, uncertainty, emotional states, licenses, learning resources, etc. At a more conceptual level we also work on modeling the Web architecture with philosophical tools so as to give a realistic account of identity and reference and to better understand the whole context of our research and its conceptual cornerstones.

3.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

One of the characteristics of Wimmics is to rely on graph formalisms unified in an abstract graph model and operators unified in an abstract graph machine to formalize and process semantic Web data, Web resources, services metadata and social Web data. In particular Corese, the core software of Wimmics, maintains and implements that abstraction. We propose algorithms to process the mixed representations of the previous section. In particular we are interested in allowing cross-enrichment between them and in exploiting the life cycle and specificity of each one to foster the life-cycles of the others. Our results all have in common to pursue analyzing and reasoning on heterogeneous semantic graphs issued from social and semantic Web applications. Many approaches emphasize the logical aspect of the problem especially because logics are close to computer languages. We defend that the graph nature of Linked Data on the Web and the large variety of types of links that compose them call for typed graphs models. We believe the relational dimension is of paramount importance in these representations and we propose to consider all these representations as fragments of a typed graph formalism directly built above the Semantic Web formalisms. Our choice of a graph based programming approach for the semantic and social Web and of a focus on one graph based formalism is also an efficient way to support interoperability, genericity, uniformity and reuse.

ZENITH Project-Team

3. Research Program

3.1. Data Management

Data management is concerned with the storage, organization, retrieval and manipulation of data of all kinds, from small and simple to very large and complex. It has become a major domain of computer science, with a large international research community and a strong industry. Continuous technology transfer from research to industry has led to the development of powerful DBMS, now at the heart of any information system, and of advanced data management capabilities in many kinds of software products (search engines, application servers, document systems, etc.).

The fundamental principle behind data management is *data independence*, which enables applications and users to deal with the data at a high conceptual level while ignoring implementation details. The relational model, by resting on a strong theory (set theory and first-order logic) to provide data independence, has revolutionized data management. The major innovation of relational DBMS has been to allow data manipulation through queries expressed in a high-level (declarative) language such as SQL. Queries can then be automatically translated into optimized query plans that take advantage of underlying access methods and indices. Many other advanced capabilities have been made possible by data independence : data and metadata modeling, schema management, consistency through integrity rules and triggers, transaction support, etc.

This data independence principle has also enabled DBMS to continuously integrate new advanced capabilities such as object and XML support and to adapt to all kinds of hardware/software platforms from very small smart devices (smart phone, PDA, smart card, etc.) to very large computers (multiprocessor, cluster, etc.) in distributed environments. For a long time, the research focus was on providing advanced database capabilities with good performance, for both transaction processing and decision support applications. And the main objective was to support all these capabilities within a single DBMS.

The problems of scientific data management (massive scale, complexity and heterogeneity) go well beyond the traditional context of DBMS. To address them, we capitalize on scientific foundations in closely related domains: distributed data management, cloud data management, big data, big data integration, scientific workflows, data analytics and search.

3.2. Distributed Data Management

To deal with the massive scale of scientific data, we exploit large-scale distributed systems, with the objective of making distribution transparent to the users and applications. Thus, we capitalize on the principles of large-scale distributed systems such as clusters, peer-to-peer (P2P) and cloud.

Data management in distributed systems has been traditionally achieved by distributed database systems which enable users to transparently access and update several databases in a network using a high-level query language (e.g. SQL) [12]. Transparency is achieved through a global schema which hides the local databases' heterogeneity. In its simplest form, a distributed database system is a centralized server that supports a global schema and implements distributed database techniques (query processing, transaction management, consistency management, etc.). This approach has proved to be effective for applications that can benefit from centralized control and full-fledge database capabilities, e.g. information systems. However, it cannot scale up to more than tens of databases.

Parallel database systems extend the distributed database approach to improve performance (transaction throughput or query response time) by exploiting database partitioning using a multiprocessor or cluster system. Although data integration systems and parallel database systems can scale up to hundreds of data sources or database partitions, they still rely on a centralized global schema and strong assumptions about the network.

In contrast, peer-to-peer (P2P) systems [8] adopt a completely decentralized approach to data sharing. By distributing data storage and processing across autonomous peers in the network, they can scale without the need for powerful servers. Popular examples of P2P systems such as Gnutella and BitTorrent have millions of users sharing petabytes of data over the Internet. Although very useful, these systems are quite simple (e.g. file sharing), support limited functions (e.g. keyword search) and use simple techniques (e.g. resource location by flooding) which have performance problems. A P2P solution is well-suited to support the collaborative nature of scientific applications as it provides scalability, dynamicity, autonomy and decentralized control. Peers can be the participants or organizations involved in collaboration and may share data and applications while keeping full control over their (local) data sources. But for very-large scale scientific data analysis, we believe cloud computing (see next section), is the right approach as it can provide virtually infinite computing, storage and networking resources. However, current cloud architectures are proprietary, ad-hoc, and may deprive users of the control of their own data. Thus, we postulate that a hybrid P2P/cloud architecture is more appropriate for scientific data management, by combining the best of both approaches. In particular, it will enable the clean integration of the users' own computational resources with different clouds.

3.3. Cloud Data Management

Cloud computing encompasses on demand, reliable services provided over the Internet (typically represented as a cloud) with easy access to virtually infinite computing, storage and networking resources. Through very simple Web interfaces and at small incremental cost, users can outsource complex tasks, such as data storage, system administration, or application deployment, to very large data centers operated by cloud providers. However, cloud computing has some drawbacks and not all applications are good candidates for being "cloudified". The major concern is w.r.t. data security and privacy, and trust in the provider (which may use not so trustful providers to operate). One earlier criticism of cloud computing was that customers get locked in proprietary clouds. It is true that most clouds are proprietary and there are no standards for cloud interoperability.

There is much more variety in cloud data than in scientific data since there are many different kinds of customers (individuals, small companies, large corporations, etc.). However, we can identify common features. Cloud data can be very large, unstructured (e.g. text-based) or semi-structured, and typically append-only (with rare updates). And cloud users and application developers may be in high numbers, but not DBMS experts.

3.4. Big Data

Big data (like its relative, data science) has become a buzz word, with different meanings depending on your perspective, e.g. 100 terabytes is big for a transaction processing system, but small for a web search engine. It is also a moving target, as shown by two landmarks in DBMS products: the Teradata database machine in the 1980's and the Oracle Exadata database machine in 2010.

Although big data has been around for a long time, it is now more important than ever. We can see overwhelming amounts of data generated by all kinds of devices, networks and programs, e.g. sensors, mobile devices, connected objects (IoT), social networks, computer simulations, satellites, radiotelescopes, etc. Storage capacity has doubled every 3 years since 1980 with prices steadily going down (e.g. 1 Gigabyte of Hard Disk Drive for: 1M\$ in 1982, 1K\$ in 1995, 0.02\$ in 2015), making it affordable to keep more data around. And massive data can produce high-value information and knowledge, which is critical for data analysis, decision support, forecasting, business intelligence, research, (data-intensive) science, etc.

The problem of big data has three main dimensions, quoted as the three big V's:

- Volume: refers to massive amounts of data, making it hard to store, manage, and analyze (big analytics);
- Velocity: refers to continuous data streams being produced, making it hard to perform online processing and analysis;
- Variety: refers to different data formats, different semantics, uncertain data, multiscale data, etc., making it hard to integrate and analyze.

There are also other V's such as: validity (is the data correct and accurate?); veracity (are the results meaningful?); volatility (how long do you need to store this data?).

Many different big data management solutions have been designed, primarily for the cloud, as cloud and big data are synergistic. They typically trade consistency for scalability, simplicity and flexibility, hence the new term Data-Intensive Scalable Computing (DISC). Examples of DISC systems include data processing frameworks (e.g. Hadoop MapReduce, Apache Spark, Pregel), file systems (e.g. Google GFS, HDFS), NoSQL systems (Google BigTable, Hbase, MongoDB), NewSQL systems (Google F1, CockroachDB, LeanXcale). In Zenith, we exploit or extend DISC technologies to fit our needs for scientific workflow management and scalable data analysis.

3.5. Data Integration

Scientists can rely on web tools to quickly share their data and/or knowledge. Therefore, when performing a given study, a scientist would typically need to access and integrate data from many data sources (including public databases). Data integration can be either physical or logical. In the former, the source data are integrated and materialized in a data warehouse. In logical integration, the integrated data are not materialized, but accessed indirectly through a global (or mediated) schema using a data integration system. These two approaches have different trade-offs, e.g. efficient analytics but only on historical data for data warehousing versus real-time access to data sources for data integration systems (e.g. web price comparators).

In both cases, to understand a data source content, metadata (data that describe the data) is crucial. Metadata can be initially provided by the data publisher to describe the data structure (e.g. schema), data semantics based on ontologies (that provide a formal representation of the domain knowledge) and other useful information about data provenance (publisher, tools, methods, etc.). Scientific metadata is very heterogeneous, in particular because of the autonomy of the underlying data sources, which leads to a large variety of models and formats. Thus, it is necessary to identify semantic correspondences between the metadata of the related data sources. This requires the matching of the heterogeneous metadata, by discovering semantic correspondences between ontologies, and the annotation of data sources using ontologies. In Zenith, we rely on semantic web techniques (e.g. RDF and SparkQL) to perform these tasks and deal with high numbers of data sources.

Scientific workflow management systems (SWfMS) are also useful for data integration. They allow scientists to describe and execute complex scientific activities, by automating data derivation processes, and supporting various functions such as provenance management, queries, reuse, etc. Some workflow activities may access or produce huge amounts of distributed data. This requires using distributed and parallel execution environments. However, existing workflow management systems have limited support for data parallelism. In Zenith, we use an algebraic approach to describe data-intensive workflows and exploit parallelism.

3.6. Data Analytics

Data analytics refers to a set of techniques to draw conclusions through data examination. It involves data mining, statistics, and data management. Data mining provides methods to discover new and useful patterns from very large datasets. These patterns may take different forms, depending on the end-user's request, such as:

- **Frequent itemsets and association rules.** In this case, the data is usually a table with a high number of rows and the data mining algorithm extracts correlations between column values. This problem was first motivated by commercial and marketing purposes (e.g. discovering frequent correlations between items bought in a shop, which could help selling more). A typical example of frequent itemset from a sensor network in a smart building would say that “in 20% rooms, the door is closed, the room is empty, and lights are on.”
- **Frequent sequential pattern extraction.** This problem is very similar to frequent itemset discovery but considering the order between. In the smart building example, a frequent sequence could say that “in 40% of rooms, lights are on at time i , the room is empty at time $i+j$ and the door is closed at time $i+j+k$ ”. Discovering frequent sequences has become critical in marketing, as well as in security (e.g.

detecting network intrusions), in web usage analysis and any domain where data come in a specific order, typically given by timestamps.

- **Clustering.** The goal of clustering is to group together similar data while ensuring that dissimilar data will not be in the same cluster. In our example of smart buildings, we could find clusters of rooms, where offices will be in one category and copy machine rooms in another because of their differences (hours of people presence, number of times lights are turned on/off, etc.).

One main problem in data analytics is to deal with data streams. Existing methods have been designed for very large data sets where complex algorithms from artificial intelligence were not efficient because of data size. However, we now must deal with data streams, sequences of data events arriving at high rate, where traditional data analytics techniques cannot complete in real-time, given the infinite data size. In order to extract knowledge from data streams, the data mining community has investigated approximation methods that could yield good result quality.

3.7. Data Search

Technologies for searching information in scientific data have relied on relational DBMS or text-based indexing methods. However, content-based information retrieval has progressed much in the last decade, with much impact on search engines. Rather than restricting the search to the use of metadata, content-based methods index, search and browse digital objects by means of signatures that describe their content. Such methods have been intensively studied in the multimedia community to allow searching massive amounts of multimedia documents that are created every day (e.g. 99% of web data are audio-visual content with very sparse metadata). Scalable content-based methods have been proposed for searching objects in large image collections or detecting copies in huge video archives. Besides multimedia contents, content-based information retrieval methods have expanded their scope to deal with more diverse data such as medical images, 3D models or even molecular data. Potential applications in scientific data management are numerous. First, to allow searching within huge collections of scientific images (earth observation, medical images, botanical images, biology images, etc.) or browsing large datasets of experimental data (e.g. multisensor data, molecular data or instrumental data). However, scalability remains a major issue, involving complex algorithms (such as similarity search, clustering or supervised retrieval), in high dimensional spaces (up to millions of dimensions) with complex metrics (L_p , Kernels, sets intersections, edit distances, etc.). Most of these algorithms have linear, quadratic or even cubic complexities so that their use at large scale is not affordable without major breakthroughs. In Zenith, we investigate the following challenges:

- **High-dimensional similarity search.** Whereas many indexing methods were designed in the last 20 years to efficiently retrieve multidimensional data with relatively small dimensions, high-dimensional data are challenged by the well-known curse of dimensionality. Only recently have some methods appeared that allow approximate Nearest Neighbors queries in sub-linear time, in particular, Locality Sensitive Hashing methods that offer new theoretical insights in high-dimensional Euclidean spaces and random projections. But there are still challenging issues such as efficient similarity search in any kernel or metric spaces, efficient construction of k-nearest neighbor graphs (k-NNG) or relational similarity queries.
- **Large-scale supervised retrieval.** Supervised retrieval aims at retrieving relevant objects in a dataset by providing some positive and/or negative training samples. Toward this goal, Support Vector Machines (SVM) offer the possibility to construct generalized, non-linear predictors in high-dimensional spaces using small training sets. The prediction time complexity of these methods is usually linear in dataset size. Allowing hyperplane similarity queries in sub-linear time is for example a challenging research issue. A symmetric problem in supervised retrieval consists in retrieving the most relevant object categories that might contain a given query object, providing huge labeled datasets (up to millions of classes and billions of objects) and very few objects per category (from 1 to 100 objects). SVM methods that are formulated as quadratic programming with cubic training time complexity and quadratic space complexity are clearly not usable. Promising solutions include hybrid supervised-unsupervised methods and supervised hashing methods.

- **Distributed content-based retrieval.** Distributed content-based retrieval methods appear as a promising solution to manage masses of data distributed over large networks, in particular when the data cannot be centralized for privacy or cost reasons, which is often the case in scientific social networks. However, current methods are limited to very simple similarity search paradigms. In Zenith, we consider more advanced distributed content-based retrieval and mining methods such as k-NNG construction, large-scale supervised retrieval or multi-source clustering.