Résumés Étendus des Thèses de Doctorat de Troisième Année 2020

Ecole Doctorale Mathématiques et Informatique

Université de Bordeaux

École doctorale Mathématiques et informatique — université de BORDEAUX

LaBRI — Institut de Mathématiques de Bordeaux — Inria — LP2N Laboratoire Photonique Numérique & Nanosciences — université de BORDEAUX

“Those who can imagine anything,
can create the impossible.”
(Alan Touring)

École Doctorale Mathématiques et Informatique ED 39

Université de Bordeaux

Année 2020

Ce recueil regroupe des résumés étendus des thèses de doctorants en troisième année à l’École Doctorale Mathématiques et Informatique de l’Université de Bordeaux. Depuis 2020, les doctorants en deuxième année de l’Ecole Doctorale Mathématiques et Informatique (EDMI) sont invités à présenter un poster dans le cadre de la journée de l’EDMI, et parallèlement les doctorants de troisième année doivent présenter un résumé.

Dans ce recueil, on trouvera 38 résumés étendus : 22 pour l’informatique et 17 pour les mathématiques. Ainsi, sans être exhaustif, ce recueil reflète le large spectre des sujets étudiés au sein de notre école couvrant les spécialités informatique, mathématiques appliquées et calcul scientifique, ainsi que mathématiques pures, et allant de la recherche fondamentale à des sujets très appliqués. Il permettra au lecteur intéressé de voir les thèmes abordés au sein des unités de recherche rattachées à notre école : IMB, LaBRI, LP2N, ou associées : INRIA, ESTIA-Recherche.

Un peu d'histoire...

1591-1991 : Quatre siècles de mathématiques à Bordeaux.

L'Université de Bordeaux fut créée le 7 juin 1441, à l'initiative de Pey Berland, cardinal-primat d'Aquitaine, par une bulle du pape Eugène IV, promulguée au concile œcuménique de Florence. Les disciplines enseignées sont la théologie, la philosophie, le droit, puis la médecine. 150 ans plus tard, le 21 juillet 1591, François de Foix-Candalle, évêque d'Ayre, captal de Buch, fonde la chaire de mathématique, et la dote de 2000 écus de pension annuelle. Son titulaire est recruté sur concours public au cours duquel il doit exposer et résoudre deux conjectures géométriques. Le prélat acquiert un immeuble pour loger et nourrir gratuitement les étudiants nécessiteux. En 1894 Jacques Hadamard rejoint à Bordeaux Pierre Duhem. Mais c'est dans les années 1960 que les mathématiques connaissent un développement important, suivies par l'informatique dans les années 70, tandis que le flux des doctorants s'accroît rapidement dans les années 80. Cette évolution conduit à la création de l'École doctorale de mathématiques en 1991 : c'est la première du genre en France.

1991-1994 : L'École doctorale de mathématiques de Bordeaux

Sous l'impulsion de ses directeurs successifs, l'École devient un pôle d'attraction : en 93-94 elle compte 112 doctorants, avec un flux moyen de 37 soutenances par an ; 8 normaliens et 1 polytechnicien y préparent leur thèse. Des conventions sont signées avec l'École Normale Supérieure de la rue d'Ulm, l'Institut Steklov et l'Université de Saint-Petersbourg. En quatre ans 17 Maîtres de Conférences, 2 Chargés de Recherches, 12 Ingénieurs sont sortis de ses rangs, et 11 doctorants étrangers obtiennent, à l'issue de leur thèse, un poste universitaire dans leur pays.

1995 : L'École doctorale de mathématiques et informatique de Bordeaux (EDMI)

L'élargissement de l'École à l'informatique, en 1995, répond à des convergences scientifiques manifestes. Beaucoup de thèmes de recherche sont voisins (Algorithmique, Parallélisme, Combinatoire, Graphisme, Imagerie, Théorie des nœuds, Dessins d'enfants de Grothendieck, Recherche Opérationnelle...). Les doctorants développent leurs recherches au sein de quatre laboratoires associés au CNRS : Laboratoire de Mathématiques Appliquées de Bordeaux - MAB (UMR 5466), Laboratoire Bordelais d'Analyse et Géométrie - LaBAG (UMR 5467), Laboratoire de Théorie des Nombres et d'Algorithme Arithmétique - A2X (UMR 5465), Laboratoire Bordelais de Recherche en Informatique - LaBRI (UMR 5800).

2007 : Création de l'IMB

Les laboratoires mathématiques – LaBAG, MAB, A2X – fusionnent en 2007 au sein de l'Institut de Mathématiques de Bordeaux (UMR 5251) créé en 2006. L'IMB devient ainsi un interlocuteur privilégié pour les mathématiques – à côté de l'INRIA installé à Bordeaux en 2008.

2014-2016 : Création de l'Université de Bordeaux et du Collège des Ecoles Doctorales

L'Université de Bordeaux est issue de la fusion entre l'Université de Bordeaux 1 (Sciences et Technologie), Bordeaux II-Ségalen (sciences de la vie, les sciences de la santé et les sciences de l'Homme) et l'Université Bordeaux IV-Montesquieu (droit, sciences économiques et sociales, sciences politiques). Un comité des écoles doctorales est créé lors de la fusion, il prend la suite d'une commission réunissant précédemment les écoles doctorales du pôle bordelais (y compris celle de l'Université Bordeaux 3, devenue depuis l'Université Bordeaux Montaigne). Ce comité est formalisé en 2016 avec la création du collège des écoles doctorales qui réunit les 8 écoles doctorales de l'Université de Bordeaux dont l'EDMI.

L'EDMI en chiffres :

Sur la période 2015-2020, l'Ecole Doctorale de Mathématiques et Informatique compte en moyenne autour de 200 doctorants dont un peu plus de la moitié en informatique et un peu moins de la moitié en mathématiques. Environ 50 thèses sont soutenues tous les ans. Ce chiffre correspond à peu près au nombre de nouveaux doctorants accueillis chaque année à l'EDMI. La durée de thèse moyenne à l'EDMI fluctue entre 40 et 41 mois. Actuellement une vingtaine de thèses se font en Convention Industrielle pour la Formation par la Recherche (CIFRE).

L'internationalisation joue un rôle fort au sein de l'EDMI ce qui est un témoin de la dynamique et représente un élément très enrichissant pour notre école. En effet, le pourcentage de doctorants d'origine étrangère est de 40%. Par ailleurs, un peu plus de vingt cotutelles avec des pays étrangers sont en cours (Algérie, Allemagne, Brésil, Canada, Chili, Chine, Espagne, Hongrie, Inde, Italie, Madagascar, Mali, Maroc, Mexique, Pays-Bas, Taïwan, Tunisie).

Les activités de l'EDMI :

Outre la gestion quotidienne des 200 doctorants,

- L'EDMI propose une offre de formation conséquente avec un spectre très large qui comprend des formations scientifiques, des cours de programmation, de la vulgarisation, etc.
- Depuis sa création, l'EDMI contribue à l'organisation et à l'essor des Leçons de Mathématiques et d'Informatique d'Aujourd'hui (LMIA). Ces leçons sont données par des représentants éminents de la communauté mathématique et informatique nationale et internationale et, depuis 2000, donnent lieu à des publications par les Éditions Cassini (Paris). Chaque exposé présente une description des motivations et notions initiales fondatrices du sujet, son évolution historique et certains aspects de son développement récent.
- Les temps forts de l'EDMI sont la rentrée solennelle à l'autonome et la journée de l'EDMI au printemps. Lors de cette journée, les doctorants sont invités à présenter leurs travaux sous forme de poster. D'autres moments forts de la journée est le concours interne « Ma Thèse en 180 secondes » de l'ED ainsi que des présentations (prix de thèse, exposé scientifique d'intérêt général, etc.).

L'EDMI entretien des relations privilégiées avec les associations des doctorants : AFoDIB en informatique et LAMBDA en mathématiques.

Direction 2020
Andreas Hartmann (directeur)
Pascal Desbarats (directeur adjoint – formation informatique)
David Lannes (directeur adjoint – formation mathématique)
Gestionnaire : Sylvaine Granier
https://ed-mi.u-bordeaux.fr/

Table des matières

[Contenu]

1 – Informatique

Optimizing preference queries in dynamic contexts

Karim ALAMI
Directeur : Sofian MAABOUT

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Preference queries present two main challenges: difficulty for users to express their preferences and the computational complexity. For skyline queries, the preferences can be on attributes, e.g., some user may look for the best flights regarding price and number of stops, and others may look for the best flights regarding number of stops and duration. In addition, preference can be expressed as a (partial) order on attributes domains, e.g., some user may prefer flight company A over B while another one may have the opposite preference. For top-k queries, users define a score function to rank objects, e.g., users who give more importance to price could define the following score function: price $*$ $2 +$ duration of the flight. In general, several rounds are required before converging towards a satisfying answer where at each round, more precise preferences are given by the user. This is due to the difficulty to figure out the precise formulation of the user's preferences. Therefore, a more or less high number of queries need to be evaluated. Our research work aims to make these queries answering faster through dedicated index structures and precomputed views. The main challenges when adopting this strategy are (i) lightweight memory consumption and (ii) fast maintenance process. Our first step was NSC, an index structure that optimizes skyline queries. However, the structure was designed for a static context making it unsuitable when data can be inserted/deleted. We redesigned NSC to cope with dynamic data and in some cases, we proposed further approaches when the structure is not suitable. In this paper, we summarize our previous contributions and present some perspective research regarding the link between regret minimization queries and what we did so far.

1 Problem Overview

Preference queries aim to retrieve points among a set of points that can be considered interesting regarding the preferences of the user over a set of parameters. They are efficient tools which reduce the amount of data returned to the user ,and consequently, avoids him an endless comparison of data.

The information retrieval regarding user preferences have historically been a ranking problem, i.e., given a set of key words, retrieve the elements that "best" match these keywords, we call it often "top-k" query . Its adaptation to relational databases has been addressed in [1]. Users are requested to put weights on the attributes in order to select points that have the higher values on attributes with the higher weights. Top-k queries gives the advantage of controlling the size of the output, however the selection of the weights remain very hard. There has been several works to simplify this process, e.g., selection of a range of weights, elicitation and regret minimizing, among others. Authors in [2] proposed the skyline operator as an alternative to ranking queries. A skyline query returns a set of points which are not dominated by any other point of the dataset. A point x dominates a point y iff x is better or equal on all attributes and strictly better on at least one attribute. The skyline query provides the advantage to not rely on a score function however, the size of the output is not controllable and it requires a quadratic computational time regarding the size of the dataset. Many studies have been done to optimize the skyline query execution time either by optimizing the number of comparisons or by materialization. For the latter, while it is highly efficient for query answering, materialized results need to be continuously updated wrt the changes in the data. In that context, we study and design efficient index structures that provide fast query answering and maintenance process.

2 Contribution

In previous work [3], the structure NSC has been presented as an index to optimize skyline queries. Its main idea consists in comparing every record r of a dataset to all remaining records r' and store the subspaces where r' dominates r in a form of pair of subspaces $compare(r, r') = \langle X|Y\rangle$ where X represents the attributes where r' is strictly better than r and Y represents the attributes where r and r' are equal. Now given a subspace Z, a record r belongs to $Sky(Z)$, i.e. the skyline over the subspace Z, if and only if there does not exist a pair $\langle X|Y\rangle$ associated to r that $covers\ Z$, i.e. $Z \subseteq XY$ and $Z \neq Y$. We denote by $cover(\langle X|Y\rangle)$ the set of subspaces covered by $\langle X|Y\rangle$. For example, consider a dataset with attributes A, B and C. The pair $\langle AB|C\rangle$ covers the subspaces $\{A, B, AB, AC, BC, ABC\}$. The time complexity of NSC is quadratic wrt the size of the dataset as well as the space complexity. However, not every pair should be kept. Let $Pairs(r)$ be the set of pairs associated to r, this set can be minimized by computing a subset $Q \subseteq Pairs(r)$ such that $cover(Q) = cover(Pairs(r))$, i.e. the set of subspaces covered by $Pairs(r)$ are covered by Q as well. We say that Q is an equivalent subset of $Pairs(r)$, $Q \equiv Pairs(r)$. The minimization problem is NP-Hard and a polynomial greedy approximate algorithm has been proposed.

In the literature, works that optimize skyline queries can be divided in three groups, works that (i) design fast algorithms without precomputation, (ii) design index structures (iii) materialize the results. We note BSKYTREE [4] the state of the art algorithm to process skyline queries without precomputing. We note HASHCUBE [5], a bitmap like index structure which associates a 2^d Boolean vector v to every record r where d is the number of attributes. $v[i]$ is set iff r belongs to $Sky(Z)$ such that Z is encoded by i. $HASHCUBE$ is highly efficient for skyline query answering, however, authors only proposed an algorithm to construct the structure from scratch in [6]. No obvious maintenance procedure is noticed. Materialized skyline views are very time and memory consuming as the number of skylines wrt to a dataset is exponential to the number of dimension. However, they provide the best query answering performance.

The experiments performed in [3] assess the construction time and memory performance of NSC against its competitors as well as query execution time. However, the structure was designed for a static context. An insertion/deletion of a record or a change in an attribute domain order may require rebuilding the structure from scratch.

For this thesis, we studied the ability of NSC to deal with updates. Precisely, we addressed its maintenance in case of (i) dynamic data, i.e., points are removed and inserted at any time and (ii) streaming data, i.e., append-only database and sliding window model. We redesigned the structure to cope with each context. Then, we addressed skyline queries optimization over a dataset with nominal attributes and dynamic order. We found that NSC is not suitable for this context, hence we established a novel approach based on views. Finally, we aim to investigate relationship between multidimensional skylines and regret minimization queries.

References

[1] S. Chaudhuri and L. Gravano. *Evaluating top-k selection queries*. In VLDB, volume 99, pages 397–410, 1999

[2] S. Börzsönyi, D. Kossmann, and K. Stocker. *The skyline operator*. In Proc. of ICDE conf., pages 421–430, 2001

[3] N. Hanusse, P. Kamnang-Wanko, and S. Maabout. *Computing and summarizing the negative skycube*. In Proc. of CIKM Conference, pages 1733–1742, 2016.

[4] J. Lee and S.-W. Hwang. *Scalable skyline computation using a balanced pivot selection technique*. Information Systems, 39:1–21, 2014

[5] K. S. Bøgh, S. Chester, D. Sidlauskas, and I. Assent. *Hashcube: A data structure for space and query efficient skycube compression*. In Proceedings of CIKM conference, pages 1767–1770, 2014.

[6] K. S. Bøgh, S. Chester, D. Sidlauskas, and I. Assent. *Template skycube algorithms for heterogeneous parallelism on multicore and GPU architectures*. In Proc. of SIGMOD Conference, pages 447–462, 2017.

Origin Equivalence for Transducers

Sougata BOSE
Directrice : Anca MUSCHOLL
Co-directeur : Gabriele PUPPIS

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

We study string transformations in the origin semantics. The origin information for a transformation takes into account how an output is produced from the input. It allows us to distinguish between same transformations that are implemented in different ways. We also develop the notion of resynchronizer, which is used to capture "similar" implementations of the same transformations and study some problems on resynchronizers, such as the synthesis problem.

1 Introduction and Motivation

Computer programs typically process input and transform it to generate output. In theoretical computer science, the input and output to a program can often be abstracted to strings, i.e, sequence of symbols from a fixed set. For example, integers can be encoded using their binary representation as a sequence of 0s and 1s. With such abstraction, studying programs essentially becomes studying transformations of strings. In many applications, the objective is to perform these transformations using as little resource as possible.

In this thesis, we study transformations of strings with restrictions on the memory used. We study the model of transducers, where the system has access to only a bounded amount of memory. The transducer stores the input and produces the output one symbol at a time. At each step of the computation, the transducer is allowed to update its internal state. As an example, consider the transformation that deletes vowels from a word. The transducer will scan the input letter by letter. The transducer stores in its internal states whether the current letter is a "vowel" or "not a vowel". Based on the state, the transducer copies the current letter only if the internal state is "not a vowel". On the input "LABRI", the transducer will be in the "not a vowel" state after reading "L", "B", and "R", and therefore, the output will be the string "LBR".

In general, there can be multiple transducers defining the same transformation. Therefore, one of the natural questions that arises in the study of transducers is the equivalence problem. Given two transducers, the equivalence problem is to determine whether they define the same transformation or not. This is a well studied problem and it was shown to be unsolvable in general[1].

However, there are some subclasses of transducers for which there are algorithms to solve the equivalence problem. One such subclass is the class of transducers that define a function, i.e, transform every input to a unique output. This is a very natural subclass that arises in many applications, for example, the study of string-to-string functions. The equivalence problem for functional transducers is decidable[2].

However, in many applications, transducers define more than just a transformation of input to output. In the example considered previously, each letter in the output "LBR" corresponds to a letter in the output "LABRI", namely, the "L", "B", and "R". This is called the origin information and it formalizes for every output letter, the input letter that is responsible for producing it. With this added origin information, a transducer defines a transformation of strings along with the origin relation between input and output letters. Thus the origin information not only captures the transformation, but also takes into account how the transformation is produced. This was introduced by Mikolaj Bojanczyk to obtain machine-independent characterization for several classes of transducers[3]. This allows us to distinguish between more transducers. Consider the transformation that takes as input a sequence of 1s and deletes one occurrence of 1. This corresponds to

computing the predecessor in the unary representation. This transformation can be implemented by either deleting the first occurrence of 1 or the last occurrence of 1. This gives us two transducers with different origins. The origin-equivalence problem asks whether two transducers define the same transformation with the same origins. In the origin-equivalence, we are able to distinguish the above two transducers, which we were not able to do in the classical equivalence.

2 Contributions

We show that the origin-equivalence problem is decidable. We obtain complexity bounds of PSPACE-complete for the origin-equivalence problem, i.e, the problem can be solved using memory that is polynomial in size of the given transducers. Moreover, this bound on the memory is optimal, i.e, all algorithms that solve this problem use at least polynomial amount of memory.

We consider some other models of transducers scan the input in a streaming manner, i.e, one symbol at a time and uses registers to produce outputs, called Streaming String Transducers. We also give algorithms to determine origin-equivalence for them. However, in this case, the bound on the memory required is not tight. Our algorithm uses memory exponential in the size of the given transducers. However, we are only able to show that any algorithm solving this problem requires polynomial amount of memory. This leaves an open problem of finding an algorithm that is optimal in terms of memory used.

We define a notion of resynchronizers to formalize the distortions of origins. Given two transducer that define the same transformation, but with different origins, a resynchronizer defines how similar the origins are. This allows us to compares transducers for equivalence up to "similar" origins. This gives rise to a notion of equivalence which is somewhere between classical and origin notions of equivalence. The resynchronizer specifies how origins of an output position is related in the two transducers. This is specified in a logical way, in the monadic second-order logic over words. We show that comparing transducers for equivalence up to resynchronizers can be done by adapting the algorithm for origin-equivalence.

We also consider the problem of synthesizing a resynchronizer. Given two transducer that define the same transformation with different origins, synthesizing a resynchronizer amounts to computing how the origins differ in the two. We show that this is undecidable in general and give algorithms in some restricted cases, in particular, the cases where the transformations are functions.

We also consider the problem of left-to-right resynchronizability. Given a transducer, we want to determine if we can define the transformation using just one left-to-right scan. This will allow us to implement the transformation in a streaming manner, i.e, the input need not be stored and the transducers will process only one input letter at a time. Given a transducer, the left-to-right resynchronizability problem asks whether it is similar to a left-to-right transducer up to some resynchronizer. This allows us to change the transformation slightly to obtain a streamable transformation which does not store the input, thus saving on the amount of memory required to implement the transformation. We give an algorithm to solve the left-to-right resynchronizability problem.

References

[1] T. V. Griffiths,*The unsolvability of the equivalence problem for lambda-free nondeterministic generalized machines, (J. ACM, 1968)*

[2] Karel Culik, Juhani Karhumäki, *The equivalence problem for single-valued two-way transducers is decidable (SIAM Journal on Computing,1987)*

[3] Mikolaj Bojanczyk, *Transducers with origin information (ICALP,2014)*

Instant deformation of surfaces in contact

Camille BRUNEL
Directeur : Pascal BARLA
Co-encadrants : Pierre BENARD, Gaël GUENNEBAUD

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux
Inria Bordeaux Sud-Ouest

In most animated films, contacts between objects are either resolved after the animation step by using costly and hardly configurable physical simulations, or completly ignored due to live constraints. This issue particularly comes up when the character's skin collides with itself or with external objects. Our work aims to resolve these contacts and propose plausible and artistically-driven deformations, only using surface-based geometrical tools without temporal dependency.

1 Introduction

Many objects of our everyday surroundings exhibit elastic behavior when put in contact with more rigid objects, e.g., a cat walking on a pillow, a hand pressed on a window, or a soft ball bouncing on a goalpost. They most notably tend to squash inside the contact region and to bulge as their volume gets redistributed outside of it. Such squashing and bulging effects are essential to communicate plausible deformations in a variety of contexts, such as animated films, visual effects or video games. They are particularly important in character animation, for instance when artists set out to convey actions of a character on the environment or on another character (e.g., grabbing, pushing, pressing, etc). Deformations in contact are thus frequently needed in practice; yet existing computer graphics tools remain of limited control for artistic use.

The classical approach to address this problem relies on the physics of elastic objects, either by explicitly simulating their behavior (e.g., [1]), or by relying on plausible, faster approximations. The obvious advantage of these approaches is their physical accuracy, provided artists manage to find the physical parameters that yield the sought-for behavior. In practice, this requires extensive training and skills, and frequent trials and errors even for a simulation expert. Physical accuracy may even be a drawback for productions that aim for cartoony, exaggerated deformations. Besides, the major limitation of simulations in an animation context is their dependence on time: the simulation needs to be run from the first frame to the current frame of the animation every time the artist modifies a parameter or the input animation. This prevents its use at the rigging or animation stage, while artists are manipulating the 3D assets.

The alternative solutions, at the opposite end of the methodological spectrum, are manual, fully artist-controlled approaches such as those based on blend shapes, free-form [2] or pose space deformation [3]. Their main advantage is their simplicity: they provide instant feedback to artists, who are then in charge of producing compelling deformations. In practice, bulging effects remain scarce in production because the task of sculpting deformations and animating them by hand requires a significant amount of time, even for accomplished artists. Even worse, each deformation is specific to the shape of objects and their actual contacts, hence it cannot be reused in different situations.

An in-between class of solutions [4] produces time-independent deformations using a quasi-static simulation — even though distant contact deformations still depend on the path taken to the colliding state. These approaches can handle more general deformations than purely geometrical deformer, but they are much more computationally demanding due to their iterative nature. Moreover, these methods offer few control to the artist and since they are based on physical laws they cannot produce exaggerated deformations typical of cartoon animation.

2 Motivations

The main objective of my PhD is to develop a deformation tool that assists the artist by managing contacts and bulging effects in an art-directable way. A seamless integration in animation workflows requires: (1) that the tool provides instant feedback to the artist; and (2) that deformations are time-independent to allow non-linear editing. For plausible bulging effects, it is also desirable that the method preserves volume to some extent; even though artistic controls should also be possible to explore more exaggerated behaviors. Only a few deformation techniques provide both interactive and time-independent solutions. However, they also exhibit important practical limitations: they either focus on interactivity and ignore volume conservation (e.g., [5]), which often yields implausible deformations; or they provide (rather time-consuming) approximations of volume conservation and are restricted to specific configurations, such as the neighbor joints of a rigged character [6], or the deformation of a cage that approximates the elastic surface [7].

3 Result

During the first part of my PhD, we developed a new approach producing plausible bulging deformations for any pair of elastic and rigid objects in intersection. By relying entirely on geometric operations, we successfully managed to achieve instant feedback with no time dependency. The resulting deformation both looks plausible and continuous in space and time even though each frame is processed independently. Our approach opens new avenues for the animation of elastic objects in contact: editing may be performed non linearly, similar deformations may be replicated in different contexts without requiring tedious manual adjustments, and many artistically driven effects can be easily accomplished by tweaking the profile curve. A result of this first work is presented in Figure 1.

In this initial work, we made one major assumption: during the collision, we considered that one object region is elastic while the other one is fully rigid. In a second phase, we strove to extend this approach to any pair of elastic-elastic surfaces.

Another major limitation of the first approach is that the bulging deformation is the same all around the contact region between the two colliding objects. The objective would then be to give the user an intuitive tool to control anisotropic deformation with advanced art-direction while producting plausible deformation. For example, when one press their thumb on the inside of their wrist, a bulge will only appear in the direction tranversal to their arm. This bulging is, in part, the result of lateral displacement of tendons. To offer a large range of plausible deformations to the user, we cannot ignore such configuration and restrict ourselves to ideal ones assuming a homogeneous inner body structure. In a last step, we would like to explore ameliorations to produce an artistically-controllable anisotropic deformation around the contact region to be able to reproduce such effects.

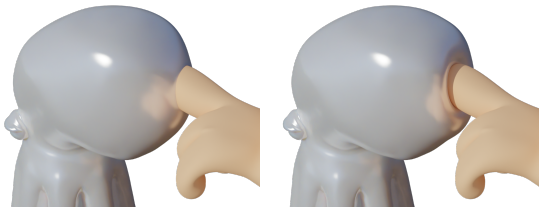

Figure 1: ***Rigid-Elastic deformation.*** *Left : a rigid finger and an elastic octopus are intersecting in the initial configuration. Right : result of our deformer, the deformed octopus skin exhibits volume preserving bulges with smooth deformation.The squid model* Teodule *is part of SOFA (*https://www.sofa-framework.org*/).*

References

[1] A. Nealen, M. Müller, R. Keiser, E. Boxerman, M. Carlson: *Physically based deformable models in computer graphics*. Computer Graphics Forum 25, 4 (2006), 809–836.

[2] T. W. Sederberg, S. R. Parry: *Free-form deformation of solid geometric models*. SIGGRAPH Comput. Graph. 20, 4 (1986), 151–160.

[3] J. P. Lewis, M. Cordner, N. Fong: *Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation*. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (2000), pp. 165–172. 1, 2

[4] A. Mc Adams, Y. Zhu, A. Selle, M. Empey, R. Tamstorf, J. Teran, E. Sifakis: *Efficient elasticity for character skinning with contact and collisions*. ACM Trans. Graph. 30, 4 (2011), 37:1–37:12.

[5] W. Wang: *A Collision Deformer for Autodesk Maya*. Master's thesis, Texas A & M University, 2015.

[6] R. Vaillant, G. Guennebaud, L. Barthe, B. Wyvill, M.-P. Cani: *Robust iso-surface tracking for interactive character skinning*. ACM Trans. Graph. 33, 6 (2014), 189:1–189:11.

[7] G. Aldrich, D. V. Pinskiy, B. Hamann: *Collision-driven volumetric deformation on the gpu*. In Eurographics (Short Papers) (2011), pp. 9–12.

AFF3CT: A Fast Forward Error Correction Toolbox

Adrien CASSAGNE
Directeurs : Christophe JÉGO & Denis BARTHOU
Co-encadrants : Camille LEROUX & Oliver AUMAGE

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux
Centre de Recherche Inria Bordeaux - Sud-Ouest, Université de Bordeaux
Laboratoire de l'Intégration du Matériau au Système, CNRS UMR 5218, Université de Bordeaux

AFF3CT is an open source toolbox dedicated to digital communication systems and specialized in channel coding. The toolbox is written in C++ and can be used either as a simulator to quickly evaluate algorithms characteristics, or as a runtime Embedded Domain Specific Language (EDSL) for the Software Defined Radio (SDR) systems design. Most of the algorithms implementations aim at low latency and high throughput, targeting multiple Gb/s on modern CPUs. This is crucial in SDR use cases: the implementations in real systems have to be very efficient to be competitive against dedicated hardware ones.

It is now commonplace to state that Humanity has entered the era of communication. The fifth generation of mobile networks (5G) is a perfect illustration of this: all kinds of objects will increasingly use communication technology to exchange information on the Internet of Things (IoT).

Despite their variety, all communication systems are based on a common abstract model proposed by Claude Shannon. In his seminal paper [1], he proposed to model a communication system with five components: an information source, a transmitter, a channel, a receiver and a destination.

Traditionally those communication systems are implemented on dedicated hardware (ASIC) targeting high throughputs, low latencies and energy efficiency. However, the implementations of such solutions have a long time to market, are expensive and are specific by nature.

The new standards like the 5G are coming with very large specifications and multiple possible configurations.Small objects that need to communicate very few data at low rates will live together with 4K video streaming for mobile phone games which will required high throughputs as well as low latencies.

To meet those various specifications the transceivers will have to be able to adapt quickly to new configurations. There is a growing need for flexible, reconfigurable and programmable solutions. To match those constraints, there is a growing interest for the SDR which consists in processing both the Physical (PHY) and Medium Access Control (MAC) layers in software, as opposed to the traditionally hardware-based solutions. Short time to market, lower development costs, interoperability, readiness to adapt to future updates and new protocols are the mains advantages of the SDR.

SDR can be implemented on various targets like Field Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs)or General Purpose Processors (GPPs).This thesis focuses on GPPs targets as they are widely spread, affordable and easy to program. More precisely, only the CPUs are considered. Many elementary blocks of digital communication systems has been optimized to run fast on x86 and ARM CPUs [5, 6, 7, 9, 11, 12] and have been packaged in AFF3CT [8, 10], our dedicated open source toolbox. Even if there is some interesting results in term of throughput on GPUs, the achieved latency on this architecture is still too high to meet real time constraints and cannot compete with existing CPU implementations.This is mainly due to data transfers between the host (CPUs) and the device (GPUs), and to the intrinsic nature of the GPU architecture which is not optimized for latency efficiency.

Some preliminary works has been made to support a part of a communication standard (DVB-S2 receiver) on CPUs and shows that it is possible to reach high throughputs (10 Gb/s) using a small cluster [2]. Those works demonstrate the feasibility but lack an overall methodology. Indeed, the approach is ad hoc for fixed parameters of the standard and a single CPU.

The Shanon's communication model presented before can be refined in a way that the transmitter and the receiver are decomposed into many processing blocks. Those blocks are mainly connected to each other in an ordered graph. This perfectly matches the dataflow model:the blocks are the filters and the binding links between the blocks represent the data movements. The dataflow model allows to describe the system from a high level point of view and to perform optimizations independently of the system designer. In the case of the SDR, simpler models than the generalized dataflow can be used like the synchronous dataflowor the cyclo-static dataflow.This allows to perform aggressive simplifications like a static scheduling of the filters execution.

Many languages dedicated to streaming applications have been developed.Streaming applications are most of the time represented with the dataflow model and the dedicated languages often support the general or the cyclo-static dataflow model. They also very often come with automatic parallelism mechanisms like pipelining and forks/joins.

However, there are few solutions targeting specifically the SDR sub-domain yet. The most famous is GNU Radio [3] which is open source and largely adopted by the community. The software is bundled with a large variety of algorithms existing in real life systems. GNU Radio models the digital communication systems at the symbol level: this philosophy is very close to the algorithms descriptions that can be found in the signal community literature. Thus, it allows quick implementation of new algorithms. Still, we show this is a limitation to meet high throughputs and low latencies constraints on current GPPs architectures, this is why, in this thesis, we propose, through AFF3CT, a runtime Embedded Domain Specific Language (EDSL) working on sets of symbols (aka frames). AFF3CT is a form of the synchronous dataflow model specialized to the relevant characteristics of FEC communication chains enabling to perform more aggressive optimizations than GNU Radio, at the cost of lower generality.

References

[1] C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal, **27**, (1948) 623–656.

[2] E. Grayver and A. Utter, *Extreme Software Defined Radio – GHz in Real Time*, Aerospace Conference (AeroConf), IEEE, 2020.

[3] *GNU Radio*, `https://www.gnuradio.org/`.

[4] P. Giard, G. Sarkis, C. Leroux, C. Thibeault and W. J. Gross, *Low-Latency Software Polar Decoders*, Springer Journal of Signal Processing Systems (JSPS), **90**, (2016) 761–775.

[5] A. Cassagne, B. Le Gal, C. Leroux, O. Aumage and D. Barthou, *An Efficient, Portable and Generic Library for Successive Cancellation Decoding of Polar Codes*, International Workshop on Languages and Compilers for Parallel Computing (LCPC), Springer, 2015.

[6] A. Cassagne, T. Tonnellier, C. Leroux, B. Le Gal, O. Aumage and D. Barthou, *Beyond Gbps Turbo decoder on multi-core CPUs*, International Symposium on Turbo Codes and Iterative Information Processing (ISTC), IEEE, 2016.

[7] A. Cassagne, O. Aumage, C. Leroux, D. Barthou and B. Le Gal, *Energy Consumption Analysis of Software Polar Decoders on Low Power Processors*, European Signal Processing Conference (EUSIPCO), IEEE, 2016.

[8] A. Cassagne, O. Hartmann, M. Léonardon, T. Tonnellier, G. Delbergue, C. Leroux, R. Tajan, B. Le Gal, C. Jégo, O. Aumage and D. Barthou, *Fast Simulation and Prototyping with AFF3CT*, International Workshop on Signal Processing Systems (SiPS), IEEE, 2017 (demo night).

[9] A. Cassagne, O. Aumage, D. Barthou, C. Leroux and C. Jégo, *MIPP: A Portable C++ SIMD Wrapper and its use for Error Correction Coding in 5G Standard*, Workshop on Programming Models for SIMD/Vector Processing (WPMVP), ACM, 2018.

[10] A. Cassagne, O. Hartmann, M. Léonardon, K. He, C. Leroux, R. Tajan, O. Aumage, D. Barthou, T. Tonnellier, V. Pignoly, B. Le Gal and C. Jégo, *AFF3CT: A Fast Forward Error Correction Toolbox!*, Elsevier SoftwareX, **10**, (2019) 100345.

[11] M. Léonardon, A. Cassagne, C. Leroux, C. Jégo, L-P. Hamelin and Y. Savaria, *Fast and Flexible Software Polar List Decoders*, Springer Journal of Signal Processing Systems (JSPS), **91**, (2019) 937–952.

[12] A. Ghaffari, M. Léonardon, A. Cassagne, C. Leroux and Y. Savaria, *Toward High Performance Implementation of 5G SCMA Algorithms*, IEEE Access **7**, (2019) 10402–10414.

Deep Curiosity: Intrinsic Motivation and Deep Learning for the Acquisition of Behavioral Repertoires in Autonomous Robotics

Cédric COLAS
Directeur : Pierre-Yves OUDEYER

INRIA Bordeaux Sud-Ouest, Flowers Team

The objective of this thesis is to develop algorithms allowing robots to learn repertoires of behaviors that are not pre-defined. The approach aims to unify two families of techniques: intrinsically motivated developmental learning and deep reinforcement learning. Learning agents are powered by intrinsic motivations, pushing them towards the acquisition and mastery of new and diverse behaviors. This learning is powered by recent advances leveraging backpropagation methods and deep neural networks for function approximation.

1 Background

A major challenge in robotics is the autonomous learning of behavioral repertoires in high-dimensional, unknown, changing environments. Autonomous learning agents must be intrinsically motivated to explore their environment, discover potential goals, represent them and learn how to achieve them. This thesis is based on contributions emerging from two families of techniques, and aims to unify them: Intrinsically Motivated Goal Exploration Processes (IMGEP) and Deep Reinforcement Learning (DRL).

1.1 Intrinsically Motivated Goal Exploration Processes

Autonomous agents must concurrently learn how to represent goals, select which ones to pursue, learn to tell whenever one is reached and learn how to achieve them. Key mechanisms in such exploration are intrinsic motivations: specific brain mechanisms that trigger spontaneous exploration for the mere purpose of experiencing novelty, surprise or learning progress [2, 10]. Algorithmic models of intrinsic motivation were successfully used in developmental robotics [3, 1]. Intrinsically Motivated Goal Exploration Processes (IMGEP) in particular, enable agents to pursue their own goals without external rewards [4, 5, 6],

1.2 Deep Reinforcement Learning

In reinforcement learning (RL) problems, an agent learns to perform a sequence of actions in an environment to maximizes some notion of cumulative reward. DRL is a family of techniques that leverage the power of deep neural networks and modern computation resources to solve such problems. DRL techniques achieved great advances in the past years such as beating humans in chess, go or video games like Atari [7]. Recently, the community has paid more attention to multi-goal and multi-task learning, where agents must learn to solve multiple tasks concurrently [8].

2 Contributions

2.1 Intrinsic Motivations for Exploration in Single Task Setting

In a first project, we proposed to perform a first stage of pure exploration using goal directed learning algorithms before leveraging that exploration for the purpose of solving a given task [12]. For each environment, we define an outcome representation that summarizes the agent's trajectory in a low-dimensional representation. A goal-directed exploration algorithm is then used to foster the exploration of that outcome space, generating diverse trajectories. These trajectories are then used to initialize the memory of a DRL algorithm, that leverages the diversity of experience to learn to solve tasks faster, and/or to avoid local optima.

2.2 Extending the Set of Tasks

In multi-goal RL, agents must learn to solve multiple tasks concurrently, where different tasks are parameterized by different *goals* or *targets* (e.g. learning to reach position X, where X belongs to a continuous space). Usually, the robots learns to achieve one type of behavior (e.g. reaching), where this behavior is modulated by various goals (e.g. reaching targets). With CURIOUS, we propose a new controller architecture to enable agents to target multiple types of goals (e.g. reach, pick and place, stack) and a continuous space of targets within each goal type [11]. This enables faster learning than having multiple policies learned in parallel. Introducing intrinsic motivations to guide behavioral acquisition, the agent first focuses on easier goals, avoids already solved ones and impossible ones, detects and corrects sensory failures.

2.3 Language as a Cognitive Tool to Imagine Goals

In previous works, the goal space is usually pre-defined by the engineer. A truly autonomous agents must learn to represent goals by itself. We introduced the IMAGINE agent, an agent able to represent goals and to learn a goal-achievement function through interactions in natural language with a social partner [13]. Through its own exploration, the agent discovers new interactions which, when relevant for the social partner, trigger a natural language description from it. Our proposed agents learn builds a set of targetable goals, learns to recognized when they are reached and trains autonomously to learn how to reach them robustly. It can also compose known goals to imagine new ones. These imagines goals foster directed exploration of its environment. The agent can further train autonomously on these imagined goals by leveraging generalization properties of its controller (it already knows how to reach them) and of its goal-achievement function (it already knows when they are reached).

References

[1] Baldassarre, G. and Mirolli, M. *Intrinsically motivated learning in natural and artificial systems.* Springer, 2013.

[2] Kaplan, F., & Oudeyer, P. Y. (2007). *In search of the neural circuits of intrinsic motivation.* Frontiers in neuroscience, 1, 17.

[3] Oudeyer, P. Y., Kaplan, F., & Hafner, V. V. (2007). *Intrinsic motivation systems for autonomous mental development.* IEEE transactions on evolutionary computation, 11(2), 265-286.

[4] Baranes, A. and Oudeyer, P.-Y. *Active learning of inverse models with intrinsically motivated goal exploration in robots.* Robotics and Autonomous Systems, 61(1):49–73,2013.

[5] Forestier, S. and Oudeyer, P. *Modular active curiosity-driven discovery of tool use.* In2016 IEEE/RSJ Inter-national Conference on Intelligent Robots and Systems(IROS), pp. 3965–3972, Oct 2016. doi: 10.1109/IROS.2016.7759584.

[6] Forestier, S., Mollard, Y., and Oudeyer, P. *Intrinsically motivated goal exploration processes with automatic curriculum learning.* CoRR, abs/1708.02190, 2017.

[7] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). *A brief survey of deep reinforcement learning.* arXiv preprint arXiv:1708.05866.

[8] Schaul, T., Horgan, D., Gregor, K., & Silver, D. (2015, June). *Universal value function approximators.* In International conference on machine learning (pp. 1312-1320).

[9] A. Baranov, K. Dyakonov, *The Feichtinger conjecture for reproducing kernels in model spaces'*, J. Geom. Anal., **21**,(2011) 276–287.

[10] Kidd, C., & Hayden, B. Y. (2015). *The psychology and neuroscience of curiosity.* Neuron, 88(3), 449-460.

[11] Colas, C., Oudeyer, P., Sigaud, O., Fournier, P., and Chetouani, M. *CURIOUS: intrinsically motivated modu-lar multi-goal reinforcement learning.* ICML 2019.

[12] Colas, C., Sigaud, O., & Oudeyer, P. Y. (2018). *GEP-PG: Decoupling exploration and exploitation in deep reinforcement learning algorithms.* ICML 2018.

[13] Colas, C., Karch, T., Lair, N., Dussoux, J. M., Moulin-Frier, C., Dominey, P. F., & Oudeyer, P. Y. (2020). *Language as a Cognitive Tool to Imagine Goals in Curiosity-Driven Exploration.* arXiv preprint arXiv:2002.09253.

Extending Physical Spaces in Spatial Augmented Reality using Projection on a Drone

Rajkumar DARBAR
Directeur : Martin HACHET
Co-directeur : Pascal GUITTON

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

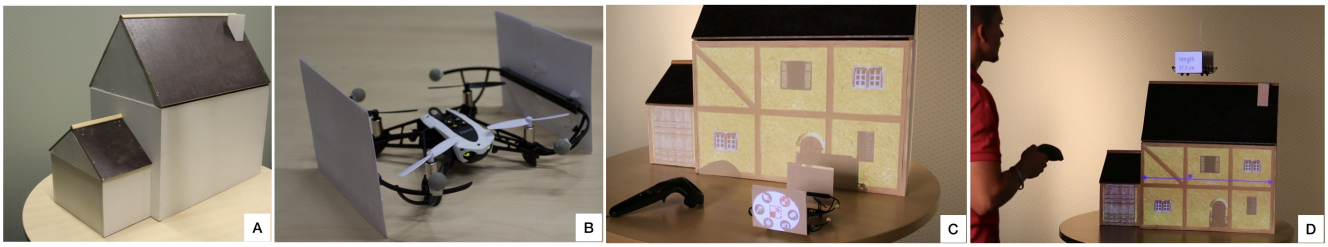

Figure 1: *An example scenario of DroneSAR. (A) A physical house mock-up. (B) A drone is mounted with two white paper panels. (C) The house is augmented using projection, and the main menu composed of a set of virtual tools projected on the drone panel. (D) A user selected the 'measuring tool' application using a controller. Then, the user positions the drone at the desired location in the 3D space (i.e., on top of the house) and draws a line shown in blue color on the augmented house to measure its width. Finally, the measured length is displayed on the drone panel.*

Spatial Augmented Reality (SAR) transforms real-world objects into interactive displays by projecting digital content using video projectors. SAR enables co-located collaboration immediately between multiple viewers without the need to wear any special glasses. Unfortunately, one major limitation of SAR is that visual content can only be projected onto its physical supports. As a result, displaying User Interfaces (UI) widgets such as menus and pop-up windows in SAR is very challenging. We are trying to address this limitation by extending SAR space in mid-air. We propose DroneSAR, which extends the physical space of SAR by projecting digital information dynamically on the tracked panels mounted on a drone. DroneSAR is a proof of concept of novel SAR User Interface (UI), which provides support for 2D widgets (i.e., label, menu, interactive tools, etc.) to enrich SAR interactive experience.

1 Introduction

Spatial Augmented Reality (SAR) [1] transforms physical surfaces into augmented surfaces by projecting digital content directly onto them. Compared to see-through augmented reality techniques, SAR allows multiple users to observe 3D augmented objects with natural depth clues, and without the need of being instrumented. This opens many opportunities in architecture [2], education [3], and so on.

Unfortunately, one of the main limitations of the SAR environment is that, contrary to see-through AR technologies, visual content can only be displayed onto physical supports. As a consequence, displaying User Interfaces (UI) widgets such as menus and pop-up windows in SAR becomes challenging. These widgets need to be positioned onto the augmented physical objects, which results in a visual clutter that affects the overall user experience. The geometry and material of the physical scene even sometimes make it impossible to

display legible UI widgets. We are trying to address these limitations by extending SAR space in mid-air. In the traditional SAR, it is not possible to display mid-air information unless using dedicated optical systems such as [5, 4] or head-tracked anamorphic illusions [6]. In this paper, we are using a flying display within the SAR environment to display mid-air content.

We propose DroneSAR, a tracked drone mounted with two rectangular white panels on which it is possible to display digital information on the fly (see Figure 1). Drones have the advantage to be flexible, as they can be positioned quickly with an acceptable accuracy around any augmented space. This allows us to extend the augmentation space and creates opportunities for new applications. In particular, DroneSAR makes it possible to embed 2D interactive widgets within the SAR experience.

The concept of extending the SAR space around the physical objects can be achieved with alternative approaches such as holding mobile devices surrounding the physical objects or adding extra projection screens around the real objects. However, our proposed solution has several benefits from its counterparts. For example, in the case of mobile devices, the users need to divide their attention between the augmented objects and the phone display. With drones, the augmentation takes place in the relevant 3D physical space, which can be at a distance from the observer. Regarding the use of extra projection screens around the objects, this makes the physical environment static, whereas the projection on a drone is more dynamic by bringing the screen where we need it. Using a robotic-arm carrying a display could be an option too, but it requires a complex motion planning setup, whereas the drones are much more flexible in terms of navigating inside a space. In our implementation, we chose to use projection rather than equipping drones with an LCD screen. This allows us to use smaller drones, which are cheaper, safer, and less noisy. Furthermore, it does not require to send synchronized video streams to the individual displays, and the rendering of the visual content remains uniform over the all augmented scene.

References

[1] Raskar, R., Welch, G., Low, K. L., & Bandyopadhyay, D. (2001). Shader lamps: Animating real objects with image-based illumination. In Rendering Techniques 2001 : Proceedings of the Eurographics, pages 89-102. DOI: *https://doi.org/10.1007/978-3-7091-6242-2_9*

[2] Tonn, C., Petzold, F., Bimber, O., Grundhofer, A., & Donath, D. (2008). Spatial augmented reality for architecture-designing and planning with and within existing buildings. International Journal of Architectural Computing, 6(1), 41-58. DOI: *https://doi.org/10.1260/147807708784640126*

[3] Laviole, J., Thevin, L., Albouys-Perrois, J., & Brock, A. (2018). Nectar: Multi-user Spatial Augmented Reality for everyone. In VRIC 2018, ACM Virtual Reality International Conference. ACM. DOI: *http://doi.acm.org/10.1145/3234253.3234317*

[4] Martinez Plasencia, D., Berthaut, F., Karnik, A., & Subramanian, S. (2014). Through the combining glass. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 341-350). ACM. DOI: *http://doi.acm.org/10.1145/2642918.2647351*

[5] Kajita, H., Koizumi, N., & Naemura, T. (2016). Skyanchor: Optical design for anchoring mid-air images onto physical objects. In Proceedings of ACM UIST 2016, pages 415-423. DOI: *http://doi.acm.org/10.1145/2984511.2984589*

[6] Benko, H., Wilson, A. D., & Zannier, F. (2014). Dyadic projected spatial augmented reality. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 645-655). ACM. DOI: *http://doi.acm.org/10.1145/2642918.2647402*

Linear/Nonlinear speech processing for analysis and classification of Parkinsonian voices

Biswajit Das
Directeur : Hussein Yahia

Institut national de recherche en informatique et en automatique, Bordeaux Sud-Ouest, Université de Bordeaux

Potential voice biomarkers are becoming increasingly desired by speech pathologists and neurologists in order to extend current noninvasive measures of speech motor abnormalities associated with neurodegeneration. Clinical information concerning acoustical features and patterns can be invaluable only if the measures are based on interpretable hypotheses and described with regard to the impact of the disease, sexual dimorphism, and any age dependency. The complexity of interpretation is the major concern between engineering applications and clinical practice. This report aims to design and define the acoustic analysis that could provide profound insight into speech disorders caused by neurodegeneration. In particular, speech feature related to each subsystem of speech production like respiration, phonation, articulation, timing and prosody, nasalic may provide important insight about localization of neurological bases. Parkinsonism refers to Parkinson's Disease (PD) and Atypical Parkinsonian Syndromes (APS), such as Progressive Supranuclear Palsy (PSP) and Multiple System Atrophy (MSA) and many others. In this study, PD, PSP and MSA disease groups are considered for differential diagnosis. The latter group (PSP and MSA) has a poor prognosis. In early stages of the disease, PD, PSP and MSA symptoms are very similar, particularly in PSP and MSA-P (parkinsonian type) patients where parkinsonism predominates. The differential diagnosis between PSP, MSA-P and PD can be very challenging in early disease stages, while early diagnostic certitude is important for the patient because of the diverging prognosis. Indeed, despite recent efforts, clinical practice demands particularly interpretable objective markers in the differential diagnosis. The need of such markers is hence very high in the neurology community, particularly given the severity of the prognosis of MSA-P. This thesis targets to develop acoustic measures to capture deficits in subsystems of speech production. Gender deformism of each developed speech markers is analyzed for accurate differential diagnosis. It also targets to find particular speech markers which are suitable for early differential diagnosis. Moreover, selection of recording task is another important parameter for measuring specific impairment. Hence, this study also analyze different recording protocols to find their suitability.

1 Introduction

Speech requires the integrity and integration of numerous neurocognitive, neuromotor, neuromuscular and musculoskeletal activities. Deficit in any of this activity may be the cause of speech disorder. It is one of the early and prominent indication for different neurological disorder. Therefore, recognizing and understanding predictable patterns of speech disturbances and their underlying neurophysiological bases are important to understand nervous system organization for speech motor control, differential diagnosis and localization of neurologic disease, prevalence and management.

1.1 State-of-the-art

Parkinson's disease (PD) is a neurological disorder caused by the degeneration of dopaminergic neurons, leading to clinical features characterized by bradykinesia, rigidity, resting tremor and postural instability.

Atypical parkinsonian syndromes (APS) such as progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) differ from PD by more widespread neuronal involvement, resulting in additional clinical signs, more rapid disease progression and poor response to dopamine replacement therapy. The majority of PSP and MSA patients develop clinical features that overlap those of PD and thus the correct diagnosis can be very challenging in early stages of the disease. In the pioneer study [1], the author stated that PSP and MSA patients manifest combination of hypokinetic, ataxic and spastic dysarthria. More precisely, MSA patients exhibit predominate ataxic dysarthria whereas PSP patients manifest hypokinetic-spastic dysarthria. PD patients exhibit pure hypokinetic dysarthria. Gender difference was also evident in the differential study [2] where male MSA-P patients manifest greater deficits in speech parameters compared to male PD patients. Study [3] showed PSP patients manifest higher deficits in articulators and laryngeal function compared to PD. In monopitch, female PSP patients manifested higher deficits compared to female PD whereas male PSP exhibited higher deficits in vowel articulation compared to male PD.

1.2 Motivation

Speech analysis has been limited to subjective auditory perceptual assessment for many generations. Subjective analysis still remain the gold standard but it is laborious task. In recent time, several studies proposed various speech markers to asses speech impairment in parkinsonism. Divergent findings in previous studies motivate us to validate previous results. Several aspects need to be explored before reaching to any conclusion about disease specific speech impairment. Gender difference, age difference, physiological conditions of patients need to be considered while differential diagnosis is targeted. To the end, development of speech dimensions (maximum by 2-dimensions) are targeted to overcome small data problem.

2 Methodology and Results

Speech database is the first requirement for any kind of analysis related to dysarthria. At the beginning, we did not have the speech database to analyze voice disorder for PD and MSA-P in French language. Hence, we started with Czech Republic Parkinsonism Database which include PD, PSP, and MSA disease groups. In parallel, we started to develop the speech database (Voice4PD-MSA) for the project Voice4PD-MSA.

With the collaborator database (Czech data), we targeted to discriminate PSP and MSA which is hardly perceptually distinguishable. In that study, we showed that PSP and MSA patients can be classified with good accuracy (88%) by linear combination of acoustic measures (dysarthric dimension). Importantly, hypokinetic and ataxic acoustic dimensions exhibited discriminative characteristics [4]. In the next study, we investigated impairment of subsystem of speech production (respiration, phonation, articulation, timing, prosody, and nasal) due to neurological disorder. This analysis revealed encouraging impairment specificity for disease groups. We developed 1-dimensional speech measure for each subsystem by simple averaging of speech dimensions. PSP patients exhibited higher deficits in respiration and particular types of articulation. In converse, MSA patients manifested predominant deficits in phonation and particular type of articulation. From clinical point of view, this observation is encouraging.

In the Voice4PD-MSA database, total 41 subjects (8 HC, 21 PD, 12 MSA-P) were recruited. Several recording task were considered to capture specific impairment. We targeted to measure deficits in vowel articulation and consonant articulation individually in details. Deficits in vowel articulation were greater for MSA-P patients compared to PD. We considered only gender independent acoustic measures to evaluate vowel impreciseness. Moreover, subjective analysis of consonants from logatomes showed greater imprecise articulators movements for MSA-P patients compared to PD.

References

[1] J. Rusz, C. Bonnet, J. Klempir, T. Tykalová, E. Baborová, M. Novotný, A. Rulseh, E. Růžička, *Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy.*, Journal of neurology. 262. 2015, 10.1007/s00415-015-7671-1.

[2] Y. E. Huh, J. Park, M. K. Suh, S. E. Lee, J. Kim, Y. Jeong, H. T. Kim, J. W. Cho, *Differences in early speech patterns between Parkinson variant of multiple system atrophy and Parkinson's disease*, Brain and Language, 147,2015, 14-20.

[3] S. Skodda, W. Visser, U. Schlegel, *Acoustical Analysis of Speech in Progressive Supranuclear Palsy*, Journal of Voice, 25, 6, 2011, 725-731.

[4] B. Das, K. Daoudi, J. Klempir and J. Rusz, *Towards Disease-specific Speech Markers for Differential Diagnosis in Parkinsonism*, ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 5846-5850.

Diffusion Vidéo avec une Meilleure Qualité d'Expérience et Respectant la Vie Privée

Simon DA SILVA
Directeurs : Daniel NÉGRU, Laurent RÉVEILLÈRE

Laboratoire Bordelais de Recherche en Informatique, LaBRI, CNRS UMR 5800, Université de Bordeaux

Le streaming vidéo devrait atteindre 82% du trafic total sur Internet en 2022. Il y a deux raisons à ce succès : la multiplication des sources de contenu vidéo et la démocratisation des connexions haut-débit à Internet. Les principales plateformes de streaming vidéo dépendent d'infrastructures planétaires pour répondre à la demande croissante en qualité visuelle. Cependant, l'utilisation de ces plateformes génère des données personnelles sensibles (sous la forme d'historiques de visionnage). Protéger les intérêts des utilisateurs est nécessaire pour une nouvelle génération de services de streaming vidéo respectueux de la vie privée. Cette thèse propose une nouvelle approche pour du streaming vidéo temps-réel multi-sources en délivrant du contenu avec une meilleure qualité d'expérience (délai de démarrage rapide, flux stable en haute qualité, pas de coupures) tout en permettant une protection de la vie privée (grâce aux environnements d'exécution de confiance).

Introduction

Le streaming vidéo représente actuellement 65% du trafic mobile sur Internet et devrait atteindre 82% du trafic total d'ici 2022. YouTube et Netflix représentent à eux seuls plus de 50% du trafic Internet en heure de pointe aux États-Unis. De plus, l'essence même de la télévision (la diffusion en direct) est en train de basculer vers des alternatives sur Internet comme Twitch, privilégiées par le public grâce à leur flexibilité d'utilisation et de support. La part du streaming vidéo en direct sur Internet est également en très forte croissance, puisqu'elle devrait être multipliée par 15 pour atteindre 17% du trafic vidéo total d'ici 2022.

Ces nouveaux modes de consommation de contenus posent un problème majeur : la *qualité d'expérience* proposée à l'utilisateur. En effet, avec une telle croissance, les opérateurs et fournisseurs de contenu peinent à mettre à niveau leurs infrastructures pour supporter la demande toujours croissante des utilisateurs. Les réseaux sont souvent saturés, les serveurs se retrouvent surchargés, et il devient de plus en plus difficile de proposer une diffusion fiable, sans coupures, avec une bonne qualité visuelle et une stabilité satisfaisante, à des coûts abordables pour les fournisseurs de contenus. Il est alors nécessaire de trouver des solutions pour réduire l'impact de ces flux sur la santé du réseau, en permettant au plus grand nombre d'accéder aux ressources tout en fournissant une bonne qualité d'expérience aux utilisateurs consommant les contenus.

Des méthodes de streaming vidéo adaptatif sur HTTP ont récemment vu le jour, notamment le standard DASH qui est utilisé entre autres par YouTube, Facebook, Netflix et Twitch. L'objectif de ces techniques est de réduire le nombre de coupures lors de la diffusion en adaptant la qualité du flux à la bande passante disponible entre l'utilisateur et le serveur. Pour cela, la vidéo est d'abord encodée dans plusieurs qualités. Ensuite, chaque qualité est découpée en segments de quelques secondes. Quand le client souhaite lire une vidéo, le serveur lui fournit une liste des qualités disponibles, et le lecteur vidéo choisit alors la qualité la plus adaptée à la bande passante disponible pour chaque segment. La vidéo est alors reçue en plusieurs segments qu'il faut remettre bout à bout pour lire le flux.

Motivation

Le standard DASH et les techniques similaires permettent d'améliorer sensiblement la *qualité d'expérience* du public en éliminant la plupart des coupures dues aux mauvaises conditions du réseau entre l'utilisateur et le serveur. En revanche, les problèmes liés à la surcharge des serveurs ou à leur capacité perdurent. Si de nombreux utilisateurs situés dans la même zone géographique regardent simultanément un même contenu vidéo, le serveur le plus proche devient rapidement surchargé. Certains utilisateurs subissent alors des dégradations de qualité ou une indisponibilité du contenu, et donc une *qualité d'expérience* faible et inéquitable.

Une autre problématique importante est la protection de la *vie privée*. L'utilisation des plateformes de streaming génère des informations personnelles sensibles en terme d'historique de visionnage aux vidéos. Ces données peuvent être exploitées soit au bénéfice de l'utilisateur, par exemple pour lui faire des recommandations personnalisées pour d'autres contenus, ou au bénéfice de la plateforme pour de la publicité ciblée. Cependant, la disponibilité des historiques d'accès peut également conduire à des menaces majeures sur la vie privée. En effet, il est facilement possible d'inférer des informations privées sur l'utilisateur, tel que son genre, origine, ses orientations politiques, religieuses ou sexuelles, ou la composition du domicile familial.

Objectif L'objectif de cette thèse est de proposer un système pragmatique et réaliste de streaming vidéo préservant la vie privée, avec à la fois une meilleure *qualité d'expérience* et des garanties de protection de la *vie privée* aux utilisateurs, au moindre coût. Proposer une bonne *qualité d'expérience* signifie (1) fournir une qualité d'image haute, (2) minimiser les fluctuations de qualité, (3) éviter les interruptions pendant la lecture, et (4) assurer un délai de démarrage rapide. Protéger la *vie privée* des utilisateurs dans un système de streaming vidéo signifie camoufler leur historique de visionnage, à la fois des serveurs et des autres utilisateurs.

Contexte Plusieurs protocoles de streaming vidéo multi-sources ont récemment émergé pour faire face à la surcharge des serveurs ou liens réseau. MS-Stream (Multiple-Source Streaming) [1], conçu au LaBRI, est un protocole de streaming vidéo adaptatif compatible avec DASH. Il permet d'utiliser plusieurs serveurs simultanément pour assurer une meilleure *qualité d'expérience* au public, à la fois en réduisant le nombre de coupures et en améliorant la qualité vidéo affichée (grâce à l'agrégation des bandes passantes).

Contributions

MUSLIN `Muslin` [2] est une solution de streaming vidéo fournissant une *qualité d'expérience* haute et équitable aux utilisateurs, à un coût inférieur aux solutions actuelles. `Muslin` implémente MS-Stream pour la livraison du contenu. `Muslin` utilise des retours périodiques automatisés des lecteurs vidéo des clients pendant les sessions de streaming, et un score de classement pour provisionner et affecter dynamiquement les serveurs selon de multiples critères. Cela permet d'ajuster l'échelle de l'infrastructure en temps réel en fonction du besoin constaté et donc réduire les coûts.

PRIVATUBE PRIVATUBE [3] est un système de streaming vidéo fournissant une bonne *qualité d'expérience* à ses utilisateurs tout en protégeant leur *vie privée*. PRIVATUBE étend MS-Stream pour améliorer la *qualité d'expérience* et réduire la charge sur les serveurs en permettant aux clients de récupérer les contenus à la fois depuis les serveurs centraux et depuis les pairs ayant regardé le même contenu précédemment. PRIVATUBE protège la *vie privée* des utilisateurs en chiffrant tous les flux dans des enclaves Intel SGX. Des requêtes fictives permettent de brouiller les pistes tout en préprovisionnant du contenu chez les pairs pour améliorer également la disponibilité et le passage à l'échelle. Nos résultats démontrent que PRIVATUBE offre un anonymat quasi-total aux utilisateurs tout en proposant une meilleure *qualité d'expérience* que les systèmes actuels.

PrivateRecSys PrivateRecSys est une solution permettant de fournir des recommandations aux utilisateurs des services de streaming tout en préservant leur *vie privée*, en garantissant un anonymat total. Pour cela, un système de double proxy dans des enclaves Intel SGX chiffre et anonymise les requêtes à la volée, de manière transparente pour les utilisateurs. Ce système est robuste aux attaques de type *side-channel*, et il a un impact faible sur la latence et le débit du service.

References

[1] J. Bruneau-Queyreix, M. Lacaud, D. Negru, J. Mongay Batalla, E. Borcoci. *Adding a new dimension to HTTP Adaptive Streaming through multiple-source capabilities*. IEEE MultiMedia 25, 3 (2018).

[2] S. Da Silva, J. Bruneau-Queyreix, M. Lacaud, D. Négru, L. Réveillère. *MUSLIN: A QoE-Aware CDN Resources Provisioning and Advertising System for Cost-Efficient Multi-Source Live Streaming*. International Journal of Network Management (IJNM '19).

[3] S. Da Silva, S. Ben Mokhtar, S. Contiu, D. Négru, L. Réveillère, E. Rivière. *PRIVATUBE: Privacy-Preserving Edge-Assisted Video Streaming*. 20th International Middleware Conference (Middleware '19).

Towards X-rays plenoptic imaging systems

Charlotte HERZOG
Directeur : Xavier GRANIER
Encadrante technique : Ombeline de La Rochefoucauld

Laboratoire Photonique, Numérique et Nanosciences, CNRS UMR 5298, Université de Bordeaux
Imagine Optic

Plenoptic is an imaging technique that allows 3D reconstruction of the scene from a single acquired image. Adapted to X-rays, this approach would allow to reduce the X-ray dose absorbed by the sample and would become a possible alternative to tomography. The work achieved in the thesis consists in a theoretical study of plenoptic optical designs extended to X-rays, a novel refocusing algorithm to reconstruct the image, and a study of depth extraction methods applied on them. A beamtime at the synchrotron PETRA III allowed us to validate this work.

1 Objective

Plenoptic or light-field imaging is a technique that captures the *light-field*, i.e. all the light rays coming from the scene. This imaging system acquires spatial and angular information of the incoming light rays separately. After the acquisition has been performed, data treatment allows image manipulation such as synthetic aperture, changing viewpoint, refocusing at different depths, and 3D reconstruction of the scene from a single acquisition.

X-ray tomography is an imaging technique that reconstructs a 3D volume from 2D projection images. It is an efficient technique that produces high resolution images. The drawback is a high irradiation of the sample, because of the high number of 2D images that needs to be acquired (usually a few thousands). Plenoptic imaging would be an alternative to tomography, as one single acquisition should be sufficient to reconstruct the volume. Achieving X-ray plenoptic imaging was the main goal of the European FET-open VOXEL project.

This thesis focused on three main parts:

- studying plenoptic theory in the visible and adapting it to X-rays, defining formulas to build and optimize a plenoptic setup
- developing an algorithm to refocus plenoptic images at various depths, as a first step to reconstruct a 3D volume
- studying Depth From Focus methods to extract depth information from the refocused images.

2 Optical system

Various systems have been designed to acquire the light-field in the visible. In this thesis, we consider plenoptic imaging systems composed of a main lens, a microlens array and a detector. Two different configurations have been presented using the same optical components: the traditional [1] and the focused [2] plenoptic setups (see Fig. 1). These configurations are usually studied separately in the literature, but they only differ in the distances between the elements in the setup.

During the thesis, these two configurations were studied, in order to highlight the pros and cons of each of them, as well as their underlying assumptions. The objective was to choose which one was the most suitable for X-rays imaging, taking into account the constraints of X-rays. We highlighted that it is possible to switch from one to the other by a continuous change in the distances, showing a continuity between the two systems [3]. We also extended the work of [1, 2] to more accurate formulas on optical configuration and theoretical resolutions.

The resolutions along the depth axis were studied in details, as depth reconstruction and extraction is

the main interest of plenoptic in this context. A specific study was done on the evolution of contrast along depth in a plenoptic system [4]. We realized that contrast decreases when moving away from a privileged depth. This is embarrassing because it can affect image reconstruction and quality of depth extraction.

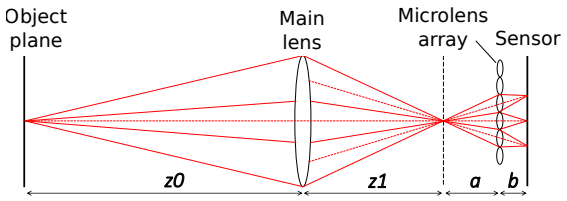

Figure 1: *Scheme of a focused plenoptic setup.*

3 Refocusing algorithms

In a second phase, we worked on refocusing algorithms. As for the optical configurations, the refocusing algorithms are usually developed for each configuration separately (traditional or focused), despite the similarities of the optical setups. The two algorithms were shown to be based on the same concept: each spatial pixel in object space is reconstructed by integrating all available angular data from the raw image [3]. An example of the result of such a refocusing process is shown in Fig. 2.

As with optical configuration, we wanted to go beyond this separation. We developed a new algorithm valid for all configurations. This algorithm is based on the same principle of integrating angular data for each spatial position. But contrary to [1, 2], our algorithm is based on the real distances between the optical elements, and needs no assumption between them.

We defined a new parameterization between object and image spaces. Using geometrical optics, we calculated the matrix transformation between the two spaces. We thus obtained equations modeling how a ray is transformed when going through the entire optical system. This allows to project data from the acquired raw image back to the object space, and reconstruct the pixels one by one, until the whole object.

Similarly, the same principle allows to project an image from object to image space. With this algorithm, we could simulate the process of image acquisition, and create synthetic plenoptic data. Reconstruction of these synthetic data was used to quantify the accuracy of the algorithm and prove its consistency.

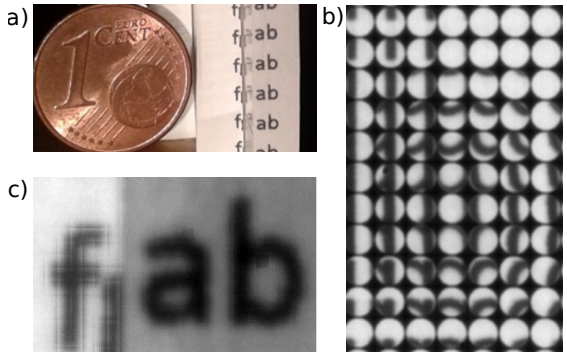

Figure 2: *Example of refocusing: a) original image with letters "ab" (size 1mm) and letters "fg" (10mm further), b) part of the raw plenoptic image (zoom on letter "b"), and c) the reconstructed image.*

4 Depth extraction

The refocusing algorithm allows to reconstruct the depth planes one by one. Each refocused plane contains the whole scene, but the elements corresponding to the current depth are sharp, whereas the ones located at other depths are blured. We used this property to extract depth from the refocused images. We tested several methods derived from the field of Depth From Focus on our images. We obtained interesting results that deserve to be deepened in some future work.

5 Application to X-rays plenoptic

In the context of the VOXEL project, the team accessed an X-rays beam at the synchrotron PETRA III. This was the occasion to apply the developed theory and algorithm to a real X-rays plenoptic experiment. In collaboration with the VOXEL partners, we were able to choose the best configuration, mount it, acquire X-rays plenoptic images, refocus them and verify some experimental resolutions. These images correspond to one of the first X-rays plenoptic images. This validates the theory and algorithm presented in the thesis. Further improvements would be done to continue this promising work.

References

[1] R. Ng, *Digital Light Field Photography*, PhD Thesis, Stanford University, (2006) AAI3219345.

[2] T. G. Georgiev, A. Lumsdaine, *Focused plenoptic camera and rendering*, Journal of electronic imaging, **19(2)**, (2010) 021106.

[3] C. Herzog, et al, *Comparison of reconstruction approaches for plenoptic imaging systems*, Proc. SPIE **10677**, Unconventional Optical Imaging, (2018).

[4] C. Herzog, et al, *Study of contrast variations with depth in focused plenoptic cameras*, Optics Letters, **44(19)**, (2019) 4825-4828.

Managing Various Application Profiles on HPC Facilities

Valentin HONORE
Directeurs : Brice GOGLIN et Guillaume PALLEZ-AUPY

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Modern High Performance Computing (HPC) offers computing facilities to solve increasingly complex scientific problems coming from many different fields. The target problems are first translated into computer programs, then run on supercomputers to obtain a solution. This thesis investigates how to optimize different types of such applications on supercomputers.

1 Context

High Performance Computing is the domain of performing complex and heavy computations on large-scale computers, called supercomputers. A supercomputer is a complex computer designed to perform until quadrillions of floating-point operations per second for the most powerful of them. A supercomputer is hierarchical: it is composed of interconnected servers, containing interconnected processors. Each processor is composed of several cores, a limited amount of memory as so as cables to interconnect all those components and the other servers. The central memory of the machine, the main storage capacities, are hard-drives located on separated boxes. Hence, this implies a really complex and dense interconnection network.

The programs that are executed on such machines are called applications. Each application aims at solving, simulating or modeling complex phenomena that a basic computer would not be able to solve due to the complexity and size of the target problem. For instance, cosmological simulations about understanding the physical behavior of dark energy and matter requires simulating environments evolving trillions of particles [1]. To do so, these simulations require storage and computing capabilities that none of personal computers are able to offer.

Nowadays, most scientific fields need supercomputers to solve their problems of interest: cosmology, physics, biology, chemistry and of course computer science. Recently, we observe a convergence between Big Data/Machine Learning and HPC. Applications of these fields (for example, Deep Learning) are becoming highly compute-intensive. Hence, HPC facilities become an appropriate solution to run such applications.

This wide range of variety of applications rise an important feature of supercomputers: they must be generic for all kind of applications. Indeed, it is hardly justified to design a machine that cost tens of million euros for only few applications able to run on it. Hence, computing nodes also have a wide range of variety, going from CPU to GPU with specific nodes designed to perform dedicated computations. Each category of node is designed to perform really fast operations of a given type (for example vector or matricial calculation).

Supercomputers are a competitive environment. Indeed, multiple users simultanously connect and request a set of computing resources to run their applications. This competition for resource is managed by the machine via a specific program called scheduler. This program assigns, referees and maps the different user requests. Each user asks for (that is, pay for the use of) resources from the supercomputer in order to run its application. Those resources are, typically, a subset of the available computational nodes with their associated features (memory and storage to store/load data etc). The resources are assigned for a limited amount of time. Competition also occurs for system resources that are shared between users, such as the bandwidth (ie, the network capacities) to access the central memory. Their availability is then hard to predict and generates side effects for applications. If too many requests arrive on the network, contention appears and requests have to wait before completion. This implies that applications

running on the machine are slowed down.

System administrators of supercomputers want their machines to be fully used at anytime, because an unused resource is a non-profitable and costly one. From user perspectives, it is important that the application runs efficiently, by fully using the capacities of requested resources. However, deciding which type of computing node to use, how many of them to request and for how long is often difficult to decide. We propose to tackle these issues in this thesis.

2 Contributions

In this thesis, we propose mathematical models, scheduling algorithms, and resource partitioning strategies in order to optimize high-throughput applications running on supercomputers. Given a target machine and application, we provide efficient scheduling strategies for minimizing total user cost of applications. In this work, we focus on two main application paradigms in the context of the convergence HPC/Big Data: data-intensive and irregular (or stochastic) applications.

Data-intensive Applications

Those applications represent typical HPC frameworks. Data-intensive applications can be decomposed into two main components. The first one is called simulation, a very compute-intensive code that generates a tremondous amount of data by simulating a physical or biological phenomena. The second component is called analytics, where sub-routines postprocess the data of simulation to extract, generate and save the final result of the application. We propose to optimize these applications on supercomputers by designing automatic resource partitioning and scheduling strategies for the two components. To do so, we use the well-known *in situ* paradigm that consists in scheduling both components together to reduce the huge cost of saving all simulation data on disks. While most of related works propose software solutions for *in situ* processing [2, 3], we propose to include into these solutions automatic resource partitioning models and scheduling heuristics to improve overall performance of *in situ* applications.

Stochastic Applications

Stochastic applications are applications for which execution time depends on its input, while in usual data-intensive applications the makespan of simulation and analytics are not affected by parameters. Stochastic jobs originate from Big Data or Machine Learning workloads, whose performance is widely dependent on characteristics of input data. Figure 1 presents an example of such applications. These applications have recently appeared on HPC platforms. However, the uncertainty of their execution time remains a strong limitation when using supercomputers. The user has to guess a first reservation value. If the job does not complete successfully within this first reservation, the user will have to resubmit the job, this time requiring a longer reservation. At the end, the total cost to the user will be the cost associated with all the reservations that were necessary to the successful completion of the job. In this thesis, we propose to model the execution time of such applications by a probability distribution (in Figure 1, we fit a LogNormal distribution) and use this knowledge to derive an optimal reservation sequence. We also derive strategies including checkpointing at the end of some (well-chosen) reservations, to avoid wasting the benefits of failed reservations. Our strategies outperforms standard approaches that consider fixed patterns of reservations regardless the execution time distribution of jobs.

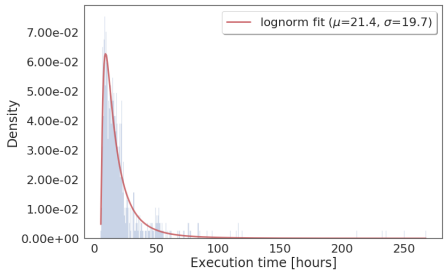

Figure 1: *Execution times from 2017 for a Structural identification of orbital anatomy application, and its fitted distribution (in red).*

References

[1] S Habib, A Pope, H Finkel, N Frontiere, K Heitmann, D Daniel, P Fasel, V Morozov, G Zagaris, T Peterka et al., *Hacc: Simulating sky surveys on state- of-the-art supercomputing architectures.*. New Astronomy, 42:49–65, Jan 2016.

[2] M Dreher and B Raffin, *A Flexible Framework for Asynchronous in Situ and in Transit Analytics for Scientific Simulations.*. 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 277-286, 2014.

[3] M Dorier, G Antoniu, F Cappello, M Snir, L Orf, *Damaris: How to Efficiently Leverage Multicore Parallelism to Achieve Scalable, Jitter-free I/O.*. CLUSTER - IEEE International Conference on Cluster Computing, Beijing, China, Sep 2012.

De l'interaction pour plus de flexibilité en fabrication additive

Sebastien IBARBOURE
Directeurs : Nadine COUTURE, Emmanuel DUC

Ecole Supérieure des Technologies Industrielles Avancées, Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Malgré leur automatisation, les procédés de fabrication additive souffrent de fréquentes erreurs de fabrication. Ces erreurs trouvent leur origine dans l'émergence de ces techniques qui aujourd'hui ne sont que partiellement comprises. Ainsi les opérateurs de fabrication traditionnellement premiers acteurs par la détection sensorielle de dysfonctionnements de fabrication, sont aujourd'hui passifs face à la fabrication additive. Nous proposons ainsi de les rendre eux aussi acteurs de cette fabrication en proposant une interface multisensorielle augmentant la fabrication.

1 Introduction

La fabrication additive (FA) est une famille de procédé qui permet la fabrication d'une entité géométrique par ajout répété de matière en couches successives. Ces procédés sont actuellement automatisés par la Chaîne Numérique, un processus industriel informatisé qui permet à partir d'une modélisation 3D de produire une pièce mécanique fonctionnelle répondant à un cahier des charges. Cette automatisation repose sur une série de transformations géométriques de ce modèle 3D pour répondre aux contraintes imposées par le procédé employé [1]. Cependant, ce processus ne garantit pas une mise en œuvre de la fabrication sans erreur dont la cause provient d'une modélisation incomplète du procédé de fabrication.

Pour améliorer la fabrication, il est donc nécessaire de détecter le plus tôt possible ces aléas pour les corriger. L'intervention de l'opérateur face à la nature imprévisible de ces aléas semble pertinente. Cependant, avec le processus actuel, aucune modification n'est possible sans la réexécution de l'ensemble du processus de fabrication et la mise au rebut de ce qui a déjà été fabriqué. De plus, le rôle de l'opérateur de fabrication additive est aujourd'hui celui d'un spectateur passif et peu expérimenté dont la mission au cours de la fabrication est réduite à une simple inspection visuelle de ce qui est fabriquée. Alors que pour les procédés de fabrication par enlèvement de matière, les opérateurs sont acteurs de la fabrication. Ils ont acquis avec le temps une expertise sensorielle leur permettant de détecter l'apparition de dysfonctionnements responsables de ces aléas.

Nous proposons ainsi de faire évoluer la mise en œuvre actuelle pour permettre l'intervention de l'opérateur afin de répondre à l'apparition d'un aléa. *Comment permettre à l'opérateur d'interagir avec le procédé de fabrication additive ?*

2 La fabrication interactive

Pour cela le concept de *fabrication interactive* proposé par Willis et al [2] constitue un modèle d'interaction avec les outils de fabrication numérique. La *fabrication interactive* y est présentée comme une extension du concept d'interaction directe en réalité virtuelle. Les outils de fabrication sont alors des interfaces devant assurer une représentation continue entre ce qui est déjà fabriqué et ce qui doit l'être, tout en permettant une action physique assistée par ordinateur sur la fabrication.

Ce concept repose sur de nouveaux outils de conception liant le réel et le virtuel. Il offre de nouveaux environnements de conceptions recourant à des interacteurs physiques, à la gestuelle et à la réalité augmentée pour pouvoir concevoir à l'échelle réelle et en 3D. L'intégration par *RoMA* [3] de ces technologies au sein d'une cellule robotisée permet la collaboration entre le robot et le concepteur. La fabrication d'un prototype

s'y fait par l'ajout de géométrie directement conçues par l'utilisateur sur une l'objet en cours de fabrication.

La fabrication interactive considère aussi l'utilisation manuelle de l'impression 3D. Cependant, le pilotage sans assistance est laborieux : il existe un décalage évident entre le résultat produit et l'intention de l'utilisateur [2]. L'amélioration de la mise en œuvre manuelle du procédé est proposée par D-Coil [4] qui s'approprie le concept d'artisan hybride (proposé par A.Zoran). À l'image d'un artisan, l'outil de fabrication est mis en mouvement par l'utilisateur tout en bénéficiant d'une assistance informatique qui en facilite la maîtrise. Cette assistance corrige localement les mouvements de l'outil pour permettre la fabrication du modèle numérique mais tout en garantissant un degré de liberté qui permet à l'utilisateur d'exprimer sa créativité. Cependant, dans D-Coil, l'extrusion de matière ne sert qu'à rendre tangible le processus de conception et n'est pas envisagé pour la fabrication d'un produit commercialisable.

3 Une fabrication augmentée

3.1 Une nouvelle Chaîne Numérique

Après avoir formalisé le processus de la Chaîne Numérique pour la fabrication additive, nous avons identifié qu'il est souhaitable d'intervenir lors de la conduite du procédé de fabrication. Pour cela, nous améliorons ce processus par l'ajout d'une boucle d'interaction en 3 étapes. La première étape consiste pour l'opérateur à détecter et identifier l'apparition d'un l'aléa. L'opérateur doit ensuite corriger cet aléa en prenant le contrôle du procédé depuis la cellule robotisée. Pour cela il peut soit en manipulant directement l'outil de fabrication soit en modifiant localement la stratégie de fabrication. Finalement, l'opérateur doit relancer la stratégie initiale de fabrication.

Par cette boucle d'interaction, nous nous intégrons dans le modèle la *fabrication interactive* bien que l'opérateur n'a pas ici un rôle de concepteur, mais celui de garantir le bon déroulement de la fabrication. Pour réaliser cette tâche, l'opérateur doit avoir la maîtrise du procédé, c'est-à-dire qu'il doit être capable d'évaluer la qualité du dépôt de matière. Nous avons ainsi établi une liste de critères de fabrication décrivant l'interaction du procédé du dépôt de matière. Pour réaliser cette tâche, l'opérateur doit avoir la maîtrise du procédé, c'est-à-dire qu'il doit être capable d'évaluer la qualité de ce qui est fabriqué.

Comment faire percevoir en temps réel ces critères pour qu'il puisse servir à l'évaluation de la fabrication ?

3.2 Une expérience multisensorielle

Pour cela, nous proposons de créer une interface multi-sensorielle permettant de recréer l'expérience sensorielle des opérateurs de fabrication par enlèvement de matière pour la fabrication additive. Ce type d'interface conjugue différentes modalités pour encoder de manière intelligible l'information en stimulations sensorielles. Nous recourrons à ce type d'interface pour communiquer les critères de fabrication nécessaires à l'évaluation par l'opérateur de la qualité du matériau imprimé.

Dans ce contexte, nous nous intéressons plus particulièrement à la traduction de ces critères par la modalité vibrotactile. Cette modalité est perçue comme une modalité pertinente de substitution pour notifier un utilisateur dans un contexte où les modalités visuelles et audio sont déjà engagées. Ce qui nous a amené à étudier les différentes techniques employées pour coder de l'information en un signal vibrotactile. Ce travail a permis d'extraire une liste de recommandations permettant d'envisager le design d'une interface de communication vibrotactile qui s'est concrétisé par une application web permettant à une personne de découvrir ces différentes techniques.

Un premier travail exploratoire a alors été entrepris pour évaluer la perception des variations de la géométrie d'un cordon de fabrication par une interface vibrotactile au poignet de l'utilisateur. Ce type d'interface semble pertinent pour notifier l'utilisateur du dépassement de plusieurs seuils d'alerte pour la hauteur et la largeur du cordon. Il est maintenant nécessaire de définir une stratégie de codage qui rendrait réalisable la production d'une ambiance tactile pour le suivi d'au moins 5 critères de fabrication à plusieurs niveaux d'alerte.

References

[1] R.Bonnard, J.Hascoet, P.Mognol, I.Stroud, *STEP-NC digital thread for additive manufacturing: data model, implementation and validation*, CIM, 2018.

[2] K.Willis, C.Xu, K.Wu, G.Levin, M.D.Gross, *Interactive fabrication: new interfaces for digital fabrication*, TEI, 2010.

[3] H.Peng, J.Briggs, C.Wang, K.Guo, Kevin and J.Kider, Joseph and S.Mueller, Stefanie and P.Baudisch, and F.Guimbretiere, *RoMA: Interactive Fabrication with Augmented Reality and a Robotic 3D Printer*, CHI, 2018.

[4] H.Peng, A.Zoran, F.Guimbretiere, *D-Coil: A Hands-on Approach to Digital 3D Models Design*, CHI, 2015.

Invariants and Algebraic structures on combinatorial objects

Théo Karaboghossian
Directeur : Jean-Christophe Aval, Adrian Tanasa

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

The research domain of my PhD is algebraic combinatorics, and in particular defining and studying algebraic structures such as Hopf algebras, species, Hopf monoids, operads and invariants - Tutte polynomial, chromatic polynomial etc - on combinatorial objects; mainly graphs, hypergraphs and similar objects.

Combinatorial invariant comparison

The first part of my research was comparing three different polynomial invariants: the one defined in [6] — which is obtained as the exponential of characteristic functions of prime elements of a Hopf algebra —, the Aguiar–Bergeron–Sottile character [3] on combinatorial Hopf algebras, and Aguiar and Ardila character on Hopf monoids defined in [1]. The first of these three invariants was introduced in the context of a certain Hopf algebra over graphs where the coproduct is defined *via* a deletion - contraction rule. This invariant gives a useful and particular elegant way to define the Tutte polynomial of a graph and to show its universality. The other two invariants are defined in similar fashions and give analogous results e. g the chromatic polynomial of graphs when considered over the canonical Hopf algebra and Hopf monoid over graphs.

In my study of these invariants I showed that the first of these invariants does not seems to have useful applications outside the particular context where it was introduced. I also obtained, anew and independently of other works, that the Hopf monoid of set compositions along the character which sends a composition P to the polynomial $\binom{t}{l(P)}$ is a terminal element in the category of combinatorial Hopf monoids (pairs of Hopf monoids and characters). And hence implies that the Aguiar–Bergeron–Sottile character is the same as Aguiar and Ardila character under the Fock functor, which sends every Hopf monoid on an associated Hopf algebra.

Chromatic polynomial of hypergraphs and reciprocity theorem

The conclusion of the preceding project came at the same time as the publication of the extensive paper on Hopf monoids [2] of Aguiar and Ardila.

The second part of my research consisted on applying the methods developed in [2] to define a polynomial character on the Hopf monoid of hypergraph also defined in [2]. This led to an interesting chromatic polynomial χ on hypergraphs in the sense that not only it generalizes the chromatic polynomial on graphs, but also Stanley reciprocity theorem on graphs . This result can be stated as follows, where acyclic orientations on hypergraphs and compatible pairs are defined in [5].

Theorem 1 (Theorem 18 and Theorem 24 in [5])**.** *Let H be a hypergraph and n an in integer. Then the number of colorings of H with $[n]$ such that each edge has only one vertex of maximal color is a polynomial on n, $\chi(H)(n)$. Furthermore, $\chi(H)(n)$ is equal to the number of compatible pairs of colorings of H and acyclic orientations on H.*

While this χ invariant was defined using the general construction of [2], the reciprocity part of the theorem was not. I used instead a more combinatorial approach by considering some functions generalizing, Faulhaber polynomials and Stirling numbers, and alterning sums over set of partitions ([5] Lemma 23). In particular the results of this lemma where first conjectured using computer investigations (using Python programming language). The facts that hypergraphs are a very general structure and that this invariant was defined using

the Hopf monoid framework also enabled me to easily define invariants, some of them already known, on other combinatorial objects: simplicial complexes, building sets etc ([5] Section 4).

Graph insertion operads

The latest part of my research was to define and study operad structures on multigraphs, hypergraphs and other similar objects and was a joint project with Samuele Giraudo. The motivations of this project are twofold. On one hand operads were successfully used on trees to define well known and rich structures such as the NAP and pre-Lie operads. On the other hand operad on graphs are used in quantum field theory to encode the combinatorics of Feynman amplitude renormalization.

Our results can be separated in two parts: a generalisation of graph operads and a study of particular cases. In the first part we gave a general way to define operads on graphs and similar objects ([4] Theorem 18). In order to do this, we introduced an isomorphism between multi-hypergraphs, i. e hypergraphs where each vertex can appear more than once in each edge, and polynomials with null constant term ([4] Remark 10). We also defined new constructions on species and operads, in particular, we defined a semi-direct product ([4] Proposition 14) which provides an operad structure on a species using an already existing operad. The second part of our project is a study of two operads of particular interest defined using the former result. These operad are a generalisation of the Kontsevich-Willwacher operad to multigraphs and a generalisation of the pre-Lie operad to oriented multigraphs. We gave an explicit link between these four operads ([4] (37)) and studied some of their finitely generated suboperads. In particular we uncovered a possible algebraic structure on some kind of pattern avoiding signed permutations and mesh patterns as well as a link with rooted forest of binary trees ([4] (50)).

References

[1] Marcelo Aguiar and Federico Ardila. The hopf monoid of generalized permutahedra. In *SIAM Discrete Mathematics Meeting*, June 2010.

[2] Marcelo Aguiar and Federico Ardila. Hopf monoids and generalized permutahedra, 2017. arXiv:1709.07504.

[3] Marcelo Aguiar, Nantel Bergeron, and Frank Sottile. Combinatorial Hopf algebras and generalized Dehn–Sommerville relations. *Compositio Mathematica*, 142(1):1–30, 2006.

[4] Jean-Christophe Aval, Samuele Giraudo, **Théo Karaboghossian**, and Adrian Tanasa. Graph operads: general construction and natural extensions of canonical operads, 2019. arXiv:1912.06563.

[5] Jean-Christophe Aval, **Théo Karaboghossian**, and Adrian Tanasa. The Hopf monoid of hypergraphs and its sub-monoids: basic invariant and reciprocity theorem, 2019. Proceeding of FPSAC 2019 (in press) arXiv:1806.08546.

[6] Gérard H. E. Duchamp, Nguyen Hoang-Nghia, Thomas Krajewski, and Adrian Tanasa. Recipe theorem for the Tutte polynomial for matroids, renormalization group-like approach. *Adv. Appl. Math.*, 51:345–358, 2013.

[7] Brittney Ellzey. A directed graph generalization of chromatic quasisymmetric functions, 2017.

[8] June Huh. Milnor numbers of projective hypersurfaces with isolated singularities. *Duke Mathematical Journal*, 163(8):1525–1548, Jun 2014.

[9] André Joyal. Une théorie combinatoire des séries formelles. *Advances in Mathematics*, 42(1):1–82, 1981.

[10] Raul Penaguiao. The kernel of chromatic quasisymmetric functions on graphs and hypergraphic polytopes, 2018.

L'Internet des Objets au service de l'Internet Media : vers une collaboration efficace

Mathias LACAUD
Directeur : Daniel NEGRU

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Les protocoles de streaming adaptatif sur HTTP sont devenue une solution de-facto pour diffuser des vidéos sur Internet grâce à leur capacité à améliorer la qualité d'expérience (QoE) des utilisateurs. Cependant, ces protocoles restent limités par la bande-passante disponible entre les clients et les serveurs situés en cœur de réseau. Au contraire, les protocoles multi-sources et pair-à-pair (P2P) offre de nouvelles opportunités d'utiliser les capacités en terme de bande-passante des équipements distribués de plus petite taille pour augmenter la stabilité et la qualité des diffusion vidéo. Nous présentons pour cela des travaux sur des protocoles hybrides P2P/multi-sources capables d'utiliser simultanément plusieurs petits serveurs et pairs pour offrir une meilleure QoE.

1 Introduction

La qualité d'expérience (*Quality of Experience* - QoE) des utilisateurs est devenu un facteur crucial pour évaluer le succès d'un système de diffusion de vidéos. Selon Cisco [3], le trafic vidéo représente 75% des données circulant sur Internet en 2020, et cette quantité de données ne va faire qu'augmenter avec l'arrivée des formats en très hautes résolutions (4K, 8K, ...). Pour faire face à une telle croissance, l'amélioration des serveurs de contenu en cœur de réseau est est une opération nécessaire. Cependant, cette opération est souvent onéreuse. Ces coûts élevés, pourtant indispensables pour garantir une bonne QoE et une bonne stabilité, posent un problème majeur pour le design des futures solutions de diffusion de vidéos. Les *Content Delivery Network* (CDNs) - des serveurs puissants et coûteux placés dans les axes réseaux stratégiques pour être proche des utilisateurs - sont actuellement massivement utilisés pour permettre des diffusions à large échelle. La plupart du temps, des techniques d'adaptation de la qualité sur HTTP (*HTTP Adaptive Streaming* - HAS), tels que les standards DASH et HLS sont utilisés pour améliorer la QoE en ajustant la qualité vidéo aux débits observés en temps réel sur le réseau. Grâce à cela, le HAS essaie d'éviter les arrêts sur image involontaires dans les vidéos (appelé "freezes"), causés la plupart du temps par une faible bande passante disponible au niveau du client ou du serveur. Même si les CDNs peuvent servir un très un grand nombre de requêtes, ils restent néanmoins contraints par la taille de leurs infrastructures physiques. À la lumière des prévisions sur l'évolution du trafic vidéo, les dépenses d'exploitation et d'investissement des distributeurs de contenus pour déployer des CDNs risquent de d'augmenter rapidement, ce qui pourrait rendre ces services coûteux pour les utilisateurs finaux en quête d'une très bonne qualité. Présentés comme une alternative aux CDNs, les systèmes pair-à-pair (P2P) pour les vidéos en direct peuvent résoudre le problème des coûts de passage à l'échelle en utilisant les ressources des utilisateurs. Cependant, ces systèmes ont aussi leurs propres limitations (volatilité et hétérogénéité des ressources par exemple) qui peuvent rapidement dégrader la QoE.

Dans le but de résoudre les problèmes de faible QoE des systèmes P2P, de manque de débit des serveurs en cœur de réseau et des importants coûts de mise à l'échelle, nous proposons d'exploiter les capacités d'équipements distribués au sein d'un système hybride P2P/multi-sources pragmatique. Dérivé des travaux [1] et [2], notre système met à profit des algorithmes de sélection et de synchronisation

multi-sources permettant de télécharger en simultané des segments vidéo depuis plusieurs sources et agréger ainsi les débits de plusieurs pairs, serveurs et équipements connectés pour obtenir une meilleure qualité à moindre coût.

2 Présentation et évaluation du système

2.1 Description du système

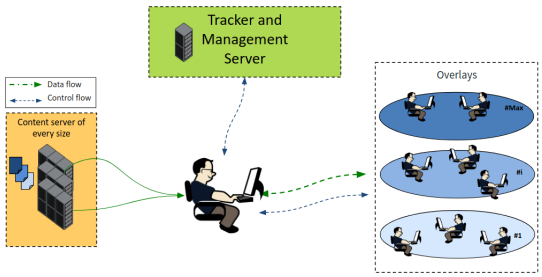

Figure 1: *Architecture du système hybride P2P/multi-sources*

Comme présenté dans la Figure 2.1, le système est composé de trois éléments : (1) des serveurs de contrôle composés d'un serveur de management qui surveille la santé du système à partir de métriques remontées pas les pairs et d'un tracker qui indique aux pairs une liste de K voisins avec qui communiquer; (2) des serveurs de toute taille, y compris des équipements distribués avec peu de ressources provisionnées avec les différentes qualités vidéo; (3) des overlays de niveau applicatif composés de N pairs consommant la même qualité et s'échangeant leurs données vidéo.

En arrivant dans le système, les pairs se placent dans un overlay, correspondant à une qualité vidéo qu'ils sont capables d'obtenir. Ils téléchargent alors une partie des données vidéo depuis les serveurs et s'échangent le reste des données vidéo entre eux. À chaque instant, les pairs remontent des informations sur le bon fonctionnement des téléchargements et des échanges au serveur de management. Ce serveur leur retourne alors des informations sur la santé globale du système. En fonction de ces informations, les pairs peuvent alors changer d'overlay pour obtenir une meilleure qualité ou changer leur comportement afin d'aider les pairs en difficulté. Tout l'enjeu est d'arriver à mettre en place des algorithmes pragmatiques et multi-critères pour que tous les pairs du système puisse avoir la meilleure qualité possible. Un exemple d'algorithmes peut être retrouvé dans [4].

2.2 Evaluation du système

Une première version de notre système de diffusion de vidéos hybride P2P/multi-sources a été implémenté en se conformant au standard DASH. Le lecteur a été développé en se servant du code de base du lecteur DASH-IF (http://dashif.org). Nous avons mené une évaluation large échelle pendant laquelle ont été consommé plus de 400 heures de vidéos avec plus de 50 utilisateurs à travers l'Europe, équipés de différents appareils (téléphones, ordinateurs avec plusieurs systèmes d'exploitation). Douze serveurs et box connectées, avec des débits hétérogènes, ont été déployés dans 7 villes de 4 pays européens.

Nous avons étudié de multiples critères considérés comme essentiels pour la mesure de la QoE, et comparé notre solution hybride P2P/multi-sources avec un lecteur standard et un lecteur P2P de l'état de l'art. Les résultats obtenus sont résumés dans le Tableau 1.

Table 1: *Résumé des résultats*

Système de diffusion de vidéos	*Classique HAS/CDN système*	*État de l'art Pair-à-pair simple*	*Notre solution hybride P2P/ multi-sources*
Qualité moyenne (doit être élevé)	3.3 Mbit/s	4.1 Mbit/s	5.9 Mbit/s
Vidéo freezes par minutes (doit être faible)	4.4	2.5	0.1
% de données récupérées par les pairs (doit être élevé)	0 %	59 %	55 %

Nous pouvons observer que notre solution permet en moyenne d'obtenir une meilleure qualité vidéo tout en diminuant le nombre d'interruptions (ou freezse) en cours de lecture. En le comparant avec un système P2P de l'état de l'art, nous pouvons constater que notre solution permet d'économiser l'utilisation des capacités du serveur d'un même ordre de grandeur d'environ 55%. Cette économie permet par exemple de servir du contenu en meilleure qualité à un plus grand nombre d'utilisateurs. Plus de détails sur ces résultats peuvent être retrouvés dans [4].

References

[1] J. Bruneau-Queyreix, M. Lacaud et al. "QoE Enhancement Through Cost-Effective Adaptation Decision Process for Multiple-Server Streaming over HTTP," in ICME. 2017.

[2] S. Da Silva, M. Lacaud et al. "MUSLIN: Achieving High, Fairly Shared QoE Through Multi-Source Live Streaming," in PV. 2018.

[3] Cisco. Cisco Annual Internet Report (2018–2023) White Paper. *http://cisco.com*

[4] J. Bruneau-Queyreix, M. Lacaud et al. "PMS: A Novel Scale-Adaptive and Quality-Adaptive Hybrid P2P/Multi-Source Solution for Live Streaming," in ACM TOMM. 2018.

Calcul de plus courts chemins multicritères

Antonin LENTZ
Directeurs : Nicolas HANUSSE, David ILCINKAS

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Le calcul de plus court chemin dans un graphe statique pondéré avec des poids réels est un problème classique pour lequel il existe des solutions satisfaisantes. En revanche, la généralisation multicritère se heurte à une complexification importante qui ne permet pas à l'heure actuelle des temps de calcul raisonnables sur des graphes ayant des millions de sommets. On s'intéresse donc à des méthodes d'approximations afin d'accélérer le temps de calcul.

1 Plus court chemin unicritère

Lorsque l'on se déplace dans un réseau de transports, l'optimisation du temps de trajet est une question importante, si ce n'est cruciale. Pour y répondre, on modélise notre réseau de tranport par un graphe pondéré, les sommets étant des lieux, reliés par des arcs dont le poids est le temps de trajet d'un lieu à l'autre. Cette modélisation a l'avantage de factoriser nombre de problèmes similaires : si l'on veut minimiser la distance parcourue, ou encore le coût financier. Plus encore, on ne se limite pas à des réseaux de transports, ces mêmes questions étant tout autant pertinentes pour les réseaux de communications à propos de la latence ou du coût.

Plus formellement, on définit un graphe orienté pondéré $G = (V, A)$ avec n sommets et m arcs. On se donne $w : A \mapsto \mathbb{R}_+$ une fonction de poids sur les arcs, ainsi que $s, t \in V$ une source et une destination. Le poids d'un chemin est la somme des poids des arêtes le composant. Le problème du plus court chemin consiste à trouver un chemin à poids minimal de s à t. Notons que ce chemin n'est pas nécessairement unique, mais on se restreint à n'en chercher qu'un.

Ce problème s'est vu proposé nombre de solutions dont les deux plus célèbres étant les algorithmes de Dijkstra dont la complexité temporelle est en $O(m + n \log n)$, et de Bellman-Ford en $O(mn)$. Enfin, l'introduction de méthodes de pré-calculs permet actuellement de répondre en quelques micro secondes à des requêtes lorsque n est de l'ordre de 10^7.

2 Multicritères

On a pu voir précedemment que plusieurs critères peuvent nous intéresser. On ajoutera à ceux déjà évoqués l'effort physique dans le cadre d'un déplacement à vélo ou à pied, ou encore le rejet en CO_2 qui mériterait une meilleure visibilité dans les outils actuels.

On peut naturellement chercher à prendre en compte ces différents critères en même temps. Cependant, il n'existe souvent pas de chemin minimisant chaque critère à la fois. Par exemple, un trajet à pied sera lent mais gratuit tandis qu'avec un taxi, il sera rapide et cher. Lequel prendre ?

Si on se donne d critères, les poids sont maintenant des vecteurs d-dimensionnels, chaque dimension correspondant à un critère. De même qu'en unicritère, si plusieurs chemins ayant même source et destination ont le même vecteur de poids, un seul nous intéresse. C'est pourquoi on assimilera un chemin à son poids. On dit qu'un chemin P domine un chemin Q si toutes ses coordonnées sont plus petites que celles de Q. Le front de Pareto d'un ensemble de chemins $\mathcal{R}$ est le sous ensemble maximal $\mathcal{S} \subseteq \mathcal{R}$ tel qu'aucun chemin de $\mathcal{S}$ ne soit dominé par un autre chemin de $\mathcal{R}$. Par exemple, sur la Fig. 1, le point G est dominé par le point B, tandis que A est incomparable avec B. De plus, les points ronds sont le front de Pareto de l'ensemble des points bleus.

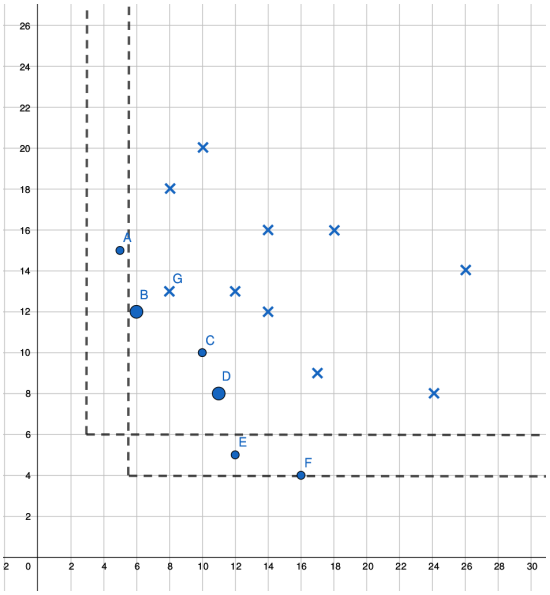

Figure 1: *Front de Pareto et approximation*

Le problème du plus court chemin devient donc, dans le cadre multicritère, la recherche du front de Pareto d'une source à une destination données. Pour se faire, les approches unicritères (Dijkstra [1] et Bellman-Ford [3]) ont été généralisées, ainsi que plusieurs améliorations classiques. Hélas, la taille du front de Pareto peut être exponentielle en le nombre de sommet, même pour $d = 2$, ce qui implique à la fois un temps de calcul exponentiel et une sortie bien trop grande pour l'utilisateur.

3 Approximation

On va donc chercher à résumer les fronts de Pareto. Pour cela, on dit qu'un chemin P de poids $(P_i)_i$ est $(1+\varepsilon)$-couvert par un chemin Q de poids (Q_i) si $\forall i, Q_i \leq (1+\varepsilon)P_i$. Un $(1+\varepsilon)$-front de Pareto d'un ensemble de chemins $\mathcal{R}$ est un sous-ensemble $\mathcal{S} \subseteq \mathcal{R}$ de chemins incomparables tel que tout chemin de $\mathcal{R}$ soit $(1+\varepsilon)$-couvert par un chemin de $\mathcal{S}$. Pour revenir à la Fig. 1, les pointillés délimitent les zones de 2-couverture de B et D, illustrant que ces deux points suffisent à former une 2-couverture de l'ensemble des points.

L'interêt majeur de cette notion se retrouve dans un résultat de [4], qui en essence nous dit qu'il existe un $(1+\epsilon)$-front de Pareto de taille de polynomiale, calculable facilement à partir d'un front de Pareto, ce qui permet de donner à un utilisateur un nombre de chemins raisonnable. Cependant, pour réduire le temps de calcul, il faudrait plutôt élaguer durant l'algorithme de découverte des chemins et non à la fin.

De même que pour le calcul exact, les algorithmes de Dijkstra [5] et Bellman-Ford [6]) sont dérivés, avec des complexités temporelles en $O\left(n^3\left(\frac{n\log(nC)}{\varepsilon}\right)^{2(d-1)}\right)$ et $O\left(nm\left(\frac{n\log(nC)}{\varepsilon}\right)^{d-1}\right)$. Cependant ces derniers sont susceptibles de renvoyer des chemins non optimaux. On ne peut donc pas comparer la taille de la sortie avec la taille du front de Pareto.

Pour le cas $d = 2$, on propose un algorithme d'approximation qui ne conserve que des chemins optimaux et dont la complexité est dans le pire cas, du même ordre de grandeur que celle de l'algorithme exact de Dijkstra. Des expériences permettent d'observer un gain non négligeable, mais qui ne devient intéressant que lorsque les front de Pareto sont très gros, cas naturellement atteint lorsque d augmente. Il convient donc de chercher à généraliser ces méthodes aux dimensions supérieures.

References

[1] P. Hansen, *Bicriterion Path Problems*, Multiple Criteria Decision Making Theory and Application, 1980.

[2] E. Martins, *On a multicriteria shortest path problem*, European Journal of Operational Research, 1984.

[3] H. Corley, I. Moon, *Shortest paths in networks with vector weights*, Journal of Optimization Theory and Applications, 1985.

[4] C. Papadimitriou, M. Yannakakis, *On the approximability of trade-offs and optimal access of Web sources*, Proceedings 41st Annual Symposium on Foundations of Computer Science, 2000.

[5] T. Breugem, T. Dollevoet, W. Van den Heuvel, *Analysis of FPTASes for the multi-objective shortest path problem*, Computers & Operations Research, 2017.

[6] G. Tsaggouris, C. Zaroliagis, *Multiobjective Optimization: Improved FPTAS for Shortest Paths and Non-Linear Objectives with Applications*, Theory Comput Syst, 2009.

Fine-grained action detection and classification from videos with spatio-temporal convolutional neural networks. Application to Table Tennis.

Pierre-Etienne MARTIN
Directrice : Jenny BENOIS-PINEAU

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Human action recognition in videos is one of the key problems in visual data interpretation. Despite intensive research, the recognition of actions with low inter-class variability remains a challenge. To answer this problem, my thesis focus on fine-grained classification challenge using a Twin Spatio-Temporal Convolutional Neural Network and apply it to a new dataset we have introduced "TTStroke-21". Our model take as input data RGB images and Optical Flow and is able to reach an accuracy of 91.4% against 43.1% for our baseline on temporal segmented videos. Detection and classification in videos using a sliding temporal window leads to a score of 81.3% over the whole dataset.

1 Introduction

The target application of our research is fine grained action recognition in sports with the aim of improving athletes' performances. Without loss of generality, we are interested in recognition of strokes in table tennis. In `TTStroke-21`[1], twenty stroke classes and an additional rejection class are considered according to the rules of table tennis. This taxonomy was designed with professional table tennis teachers. We are working on videos recorded at the Faculty of Sports of the University of Bordeaux - STAPS. Students are the athletes filmed and the teachers supervise exercises conducted during the recording sessions. The recordings are markerless and allow players to perform in natural conditions. The goal is to develop an automatic analysis tool that teachers and students can use to analyse tennis table players games to improve their performances. This dataset is the first step of my thesis.

The second step is the classification process. A new Twin Spatio-Temporal Convolutional Neural Network - TSTCNN - is introduced for this purpose [1]. Our model similarly processes RGB images and Optical Flow through a succession of spatio-temporal convolutions. A middle fusion is done before the calculation of the class scores. We compare the performances using our dataset with the baseline Two-Stream I3D method proposed in [2]. A temporal segmentation of table tennis strokes in videos is also be performed, based on temporal sliding windows and our TSTCNN classifier.

2 TTStroke-21

`TTStroke-21` is constituted of player-centred videos using GoPro cameras with 120 frames per second recorded in natural conditions. Sequences have been recorded indoors using artificial light. Experts in Table Tennis annotate the videos through an annotation platform using twenty stroke classes accordingly to the table tennis rules. To obtain an exploitable dataset, annotations had to be processed by different filters to remove annotation errors and annotation where joined when part of videos have been annotated twice by different annotators. A total of 1058 annotations are then kept and a rejection class is built upon them.

3 Method

Classification of actions is performed for a single table tennis player performing a series of strokes. Full HD video frames are resized to 320×180 pixels and their Optical Flow (OF) is computed offline.

3.1 Optical Flow and Region of Interest

Different Optical flow methods were investigated [3]. A region of interest (ROI) of size (W, H) is then inferred from the center of mass of the foreground motion amplitude map, see Fig. 1.

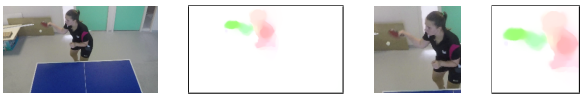

Figure 1: *ROI extraction based on optical flow*

3.2 Architecture

Our TSTCNN is constituted of 2 individual branches with three 3D convolutional layers with 30, 60, 80 filter response maps, followed by a fully connected layer of size 500, see Fig. 2. One branch takes RGB values as input, and input of the other branch is the OF preliminary estimated on the current video frame. The 3D convolutional layers use $3 \times 3 \times 3$ space-time filters with a dense stride and padding set to 1 in each direction. The two branches are fused through a final fully connected layer of size 21 followed by a Softmax function for outputting a classification score.

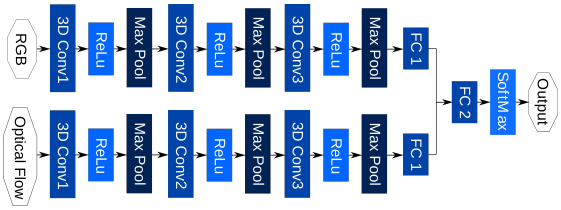

Figure 2: *TSTCNN - Twin Spatio-Temporal Convolutional Neural Network*

3.3 Model Training

The '`Twin`' model uses the full architecture presented in section 3.2 while '`RGB`' and '`Optical Flow`' models are constituted of one branch only. The optimization method is Stochastic Gradient Descent with Nesterov momentum.

3.4 Data Augmentation

Data augmentation is performed on the fly to save storage space. For spatial augmentation we apply random rotation, a random translation, and a random homothety both on RGB images and optical flow. Transformations are applied with respect to the center of the ROI. Finally we perform horizontal flip with probability of 0.5. For temporal augmentation we extract T successive frames following a normal distribution around the center of our stroke.

4 Experiments & Results

To compare the performances of our models, we use the Two-Stream I3D model introduced [2] as our baseline and apply it to our dataset following their instructions for training. The RGB images and Optical Flow streams are trained separately and a late fusion by addition of the class scores is performed to classify the action. Our Twin model perform the best reaching 91.4% against 81.9% for '`RGB`', 44.5% for '`Optical Flow`' and 43.1 for '`I3D`' using both stream.

In [3] we improve these results by changing the normalization method of the OF and perform both detection and classification simultaneously using a sliding temporal window on the whole dataset leading to a score of 81.3%.

5 Conclusion & Perspectives

My thesis aims to improve athletes performances by developing new methods and tools for coaches and students. For now on, we are able to segment temporally and classify games of Table Tennis players. In a near future, other information will be extracted in the method framework, such as characterizing the quality of a performed stroke, or establishing player statistics. The same protocol could be extended to other sports by adapting classes and size of our video cuboids according to the specific rules of the sport of interest. However, it requires to have a dataset dedicated for each sport. In our case, the `TTStroke-21` dataset, is still enriched to improve the performances of our TSTCNN model and to make possible better tools for coaches and students.

References

[1] P-E. Martin, J. Benois-Pineau, R. Péteri and J. Morlier, *Sport Action Recognition with Twin Spatio-Temporal CNNs: Application to Table Tennis*, CBMI, 2018.

[2] J. Carreira, A. Zisserman, *Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset*, CVPR, 2017.

[3] P-E. Martin, J. Benois-Pineau, R. Péteri and J. Morlier, *Optimal choice of motion estimation methods for fine-grained action classification with 3D Convolutional Networks*,ICIP, 2019.

Geographical data dissemination in the IoV: an efficient and secure SDN-based approach

Leo MENDIBOURE
Directrice : Francine KRIEF ; Co-encadrant : Mohamed Aymen CHALOUF

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

With the Internet of vehicles, the daily life of road users could be radically changed: higher road safety, smoother road traffic, improved comfort. To enable that, data must be efficiently shared between vehicles and devices located in a same geographical area: street, neighbourhood, city. The Software Defined Networking (SDN) technology could be an efficient way to enable this geographic data dissemination. However, there are many challenges related to the integration of SDN into vehicular networks: efficient definition of geographical areas, communication overhead and communication security. In this thesis, we address these challenges through different contributions: an innovative communication architecture, a SDN-based middleware for geographic data dissemination, a state machines-based approach to reduce the SDN delay-overhead and a scalable authentication and access control mechanism.

Background and Motivations

Every year, tens of millions of people are killed or injured in road accidents. Thus, the development of solutions to ensure the road users' safety is essential. With the advent of wireless communication networks and mobile terminals, the idea of vehicular communication networks has emerged. With vehicular networks, each vehicle would be able to inform its neighbours about road conditions, obstacles and sudden braking. Reaction time and accidents' number could be reduced.

Initially, a fully decentralised, *ad hoc*, architecture, based on direct communications between vehicles, was considered. However, such an architecture has many limitations: complex deployment of global services, limited interaction with the environment, etc. Therefore, vehicle networks have evolved towards a hybrid architecture, combining the benefits of an *ad hoc* approach (low latency) and a centralised approach (internet connection, efficient data processing). A new paradigm has emerged, the Internet of Vehicles (IoV), connecting vehicles to the Internet and their environment: smartphones, traffic lights, parking lot, etc.

With IoV, road safety applications are improved and new services become possible: road traffic management, comfort and entertainment. Many of these applications have a common requirement, the geographical dissemination of data. Indeed, for road safety (obstacle detection, emergency braking), traffic management (cooperative map creation, road traffic redirection), user comfort (parking lots availability, recharging points position) or even entertainment (Points of Interest, Augmented Reality), the information generated by vehicles is only relevant in a given geographical area: street, neighbourhood, town, etc.

Therefore, many research work in vehicular networks focus on geographical data dissemination. To disseminate data over large geographical areas, cellular networks may be used. Indeed, they provide higher bandwidth and lower packet loss than direct vehicle-to-vehicle communications and guarantee data dissemination even in areas with a small number of vehicles. Thus, cellular communications represent the best solution for non delay sensitive applications: cooperative mapping, traffic redirection, Points of Interest, etc. That is why, in this thesis we considered the use of cellular network to ensure an efficient geographical data dissemination in vehicular networks.

Due to the high mobility of vehicles and the short lifetime of communication links, geographical data dis-

semination must be highly flexible and responsive. To achieve that, the Software Defined Networking (SDN) technology could be the solution. Indeed, SDN provides a high network programmability, flexibility and a centralized vision. Thus, this approach could be an effective way to geographically disseminate data in vehicular networks. That is why, in this thesis we focused on the use of the SDN technology for the geographic dissemination of data in the IoV environment.

Challenges

Many research projects focus on the integration of the SDN technology in the IoV environment (SD-IoV). Nevertheless, important challenges must be addressed to enable an efficient geographic data dissemination:

- **definition of geographical areas of data dissemination using SDN:** geographical data dissemination should enable to transmit data to devices located within specific geographical areas. However, existing solutions are insufficiently flexible, involve significant overheads and require the deployment of routers supporting specific protocols. Thus, SDN-based solutions should be proposed to define these geographical areas;
- **overhead related to the SDN technology:** with SDN, the definition of communication paths involves exchanges between different devices: the routers and the SDN controller. Due to the high mobility of nodes in IoV, the number of exchanges may be significant. This could lead to important delays. Moreover, existing solutions aim to limit the exchanges duration but not the number of exchanges. In high-density areas, these solutions may be ineffective. Thus, a solution limiting the number of exchanges between the routers and the SDN controller must be defined;
- **lack of security in information routing:** security is a key issue for IoV, to ensure road users' safety, communications must be secured. Indeed, erroneous information could lead to dangerous decisions and accidents. Therefore, a secure communication between the different nodes (vehicles, base stations, SDN controllers, etc.) is necessary. However, SDN standards do not offer effective and scalable security solutions. Thus, an efficient and scalable authentication and access control mechanism must be designed for SD-IoV.

Contributions

The work we carried out in this thesis aimed to provide a solution to the problems identified above. Thus, we used the SDN technology to enable an efficient and secure geographical data dissemination in IoV through:

- **an improved SD-IoV communication architecture:** existing SD-IoV architectures present different limitations: control plane management, security and privacy plane security, global data processing. To address these challenges, we proposed an improved SD-IoV architecture [1];
- **an SDN-based geographical data dissemination:** an SDN-based definition of geographical areas has not been proposed so far. We proposed such a mechanism considering different parameters such as applications requirements, network load and terminal devices position [2]. Evaluations carried out tend to demonstrate the benefits of our approach compared to existing solutions;
- **a state machines-based approach to reduce the SDN delay-overhead:** overhead is due to the high number of exchanges between SDN controller and SDN devices. To limit these exchanges and to improve mobility management, we proposed a solution based on the deployment of state-machines at the routers level. The evaluation of this approach tends to show its benefits in terms of latency and communication overhead in comparison with existing solutions;
- **a blockchain-based authentication and access control solution for SD-IoV devices:** authentication and access control are a major concern for SD-IoV [3]. Current solutions have a major limitation: scalability. That is why we proposed a scalable architecture, composed of a set of geographical Blockchain networks, and Blockchain-based mechanisms, enabling an efficient authentication and access control for SD-IoV. The evaluation of this solution demonstrates its advantages in terms of latency, bandwidth and CPU usage.

References

[1] L. Mendiboure, M.A. Chalouf, F. Krief, *Towards a 5G vehicular architecture*, International Workshop on Communication Technologies for Vehicles, Springer, Cham, 2019.

[2] L. Mendiboure, M.A. Chalouf, F. Krief, *A SDN-Based Pub/Sub Middleware for Geographic Content Dissemination in Internet of Vehicles*, 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), IEEE, 2019.

[3] L. Mendiboure, M.A. Chalouf, F. Krief, *Towards a blockchain-based SD-IoV for applications authentication and trust management*, International Conference on Internet of Vehicles, Springer, 2018.

Problèmes de reconfiguration dans les graphes

Paul OUVRARD
Directeurs : Paul DORBEC et Cyril GAVOILLE. Encadrante : Marthe BONAMY

Laboratoire Bordelais de Recherche en Informatique (LaBRI), CNRS UMR 5800, Université de Bordeaux

Les problèmes de reconfiguration sont des problèmes étudiant l'espace des solutions d'un problème donné, tant d'un point de vue algorithmique que structurel. Ainsi, plusieurs questions se posent naturellement : (i) quelle est la complexité de déterminer s'il est toujours possible de transformer une solution en une autre ? (ii) Quelle est la longueur maximum d'une séquence de transformation minimum ? Ici, nous nous intéressons à la reconfiguration d'ensembles dominants d'un point de vue algorithmique.

1 Introduction

Les problèmes de reconfiguration apparaissent lorsque étant donnée une instance d'un problème, nous souhaitons obtenir une transformation étape par étape entre deux solutions du problème, de telle sorte que chaque étape intermédiaire soit également une solution du problème. Une telle transformation est appelée séquence de reconfiguration. Malheureusement, il n'est pas toujours possible de transformer une solution en une autre. Certaines sont même *gelées*, c'est-à-dire qu'elles ne peuvent pas du tout être modifiées.

L'intérêt pour les problèmes de reconfiguration n'a cessé de croître au cours de la dernière décennie. La reconfiguration de nombreux problèmes de théorie des graphes a été étudiée. Parmi ceux-ci, nous pouvons notamment citer Coloration, Ensemble Indépendant, Ensemble Dominant, Couverture Par Sommets. Pour un aperçu des résultats récents sur les problèmes de reconfiguration, le lecteur est invité à consulter les synthèses de van den Heuvel [1] et Nishimura [2].

2 Définitions

Nous nous intéressons ici à la reconfiguration d'ensembles dominants. Rappelons qu'un ensemble dominant d'un graphe $G = (V, E)$ est un sous-ensemble de sommets $D \subseteq V$ tel que chaque sommet n'appartenant pas à D possède un voisin dans D.

Visualisons un ensemble dominant matérialisé par des jetons placés sur les sommets qui le composent. Ainsi, modifier un ensemble dominant correspond à déplacer les jetons selon une certaine règle, appelée règle de reconfiguration. Trois opérations ont principalement été étudiées dans la littérature :

- *Token Addition and Removal* (TAR(k)) : on peut ajouter ou supprimer un jeton tant que le nombre total de jetons ne dépasse pas k ;
- *Token Jumping* (TJ) : on peut déplacer un jeton vers n'importe quel sommet du graphe ne possédant pas de jeton ;
- *Token Sliding* (TS) : on peut déplacer un jeton vers un voisin de son sommet actuel.

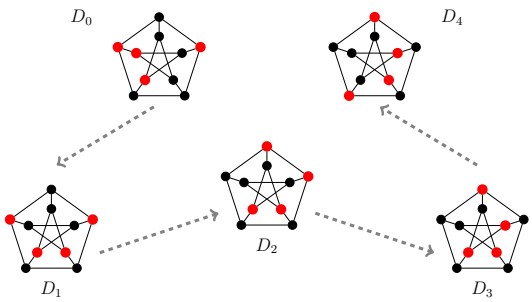

Figure 1: *Séquence de reconfiguration de D_0 à D_4 sous* TS.

3 Questions étudiées

Les principales questions étudiées en reconfiguration peuvent être exprimées en fonction du *graphe de reconfiguration* que l'on note $\mathcal{R}_G(k)$. Il est défini comme suit : les sommets de $\mathcal{R}_G(k)$ correspondent aux ensembles dominants de G de taille k (ou au plus k si l'on s'intéresse à la règle TAR(k)) et il y a une arête entre deux sommets si l'on peut passer de l'un à autre à l'aide de la règle considérée en une étape. Ainsi, trois problèmes peuvent être naturellement identifiés :

- Le problème d'*accessibilité* : existe-t-il un chemin entre deux ensembles dominants donnés D_s et D_t dans $\mathcal{R}_G(k)$?

- Le problème de *connectivité* : le graphe $\mathcal{R}_G(k)$ est-il connexe ?
- Le problème du *plus court chemin* : quelle est la distance entre deux ensembles dominants donnés D_s et D_t dans $\mathcal{R}_G(k)$?

4 Résultats connexes

La reconfiguration d'ensembles dominants a été principalement étudiée sous la règle TAR. Haas et Seyffarth ont démontré que le graphe de reconfiguration $\mathcal{R}_G(n-1)$ (où n est le nombre de sommets de G) est connexe si G possède au moins deux arêtes indépendantes [3]. La valeur du seuil peut être abaissée à $\Gamma+1$ (où Γ est la taille maximum d'un ensemble dominant minimal de G) si G est un graphe chordal ou un graphe biparti. Suzuki *et al.* ont démontré que ce résultat ne se généralisait pas à l'ensemble des graphes en exhibant une famille infinie de graphes pour laquelle $\mathcal{R}_G(\Gamma+1)$ n'est pas connexe [4].

Haddadan *et al.* ont étudié la complexité de ce problème en fonction de différentes classes de graphes [5]. Ils ont notamment prouvé que le problème d'accessibilité était PSPACE-complet, y compris si le graphe est biparti ou est un graphe scindé (*split* en anglais). D'un autre côté, ils ont construit des algorithmes s'exécutant en temps linéaire pour les graphes d'intervalles, les arbres ou encore les cographes.

Enfin, Mouawad *et al.* ont démontré que le problème est W[2]-difficile lorsque paramétré par $k+\ell$, où k est le seuil et ℓ la longueur de la séquence de reconfiguration [6]. Lokshtanov *et al.* ont construit un algorithme FPT paramétré par k pour les graphes excluant $K_{d,d}$ comme sous-graphe, pour tout constante d [7].

5 Nos résultats

Nous nous intéressons principalement à la reconfiguration d'ensembles dominants sous TS. Plus précisément, nous montrons que la question d'accessibilité est PSPACE-complète, y compris si le graphe est un graphe scindé ou un graphe biparti. D'un autre côté, nous construisons des algorithmes s'exécutant en temps polynomial pour les graphes *dually chordaux* (surclasse des graphes d'intervalles et des arbres) ou pour les cographes qui sont les graphes sans P_4 induit.

Nous nous intéressons également à la complexité de la variante optimisée suivante. Soit $G=(V,E)$ un graphe, D un ensemble dominant de G, et s,k deux entiers . Existe-t-il un ensemble dominant D' de taille au plus s et qui est accessible depuis D par la règle TAR(k) ? Ce problème est NP-difficile puisqu'il généralise Ensemble Dominant si $D=V$ et $k=|V|$. Cependant, nous prouvons qu'il est PSPACE-complet. Nous montrons également qu'il est W[2]-difficile lorsque paramétré par k. En revanche, nous exhibons des algorithmes FPT paramétrés par $d+s$ (où d est la dégénérescence de G), ou par la taille d'une couverture par sommets minimum de G.

References

[1] J. van den Heuvel. The complexity of change. In *Surveys in Combinatorics 2013*, volume 409 of *London Mathematical Society Lecture Note Series*, pages 127–160. Cambridge University Press, 2013.

[2] N. Nishimura. Introduction to reconfiguration. *Algorithms*, 11(4):52, 2018.

[3] R. Haas and K. Seyffarth. The k-dominating graph. *Graphs and Combinatorics*, 30(3):609–617, May 2014.

[4] Akira Suzuki, Amer E. Mouawad, and Naomi Nishimura. Reconfiguration of dominating sets. *Computing and Combinatorics*, pages 405–416, Cham, 2014. Springer International Publishing.

[5] Arash Haddadan, Takehiro Ito, Amer E. Mouawad, Naomi Nishimura, Hirotaka Ono, Akira Suzuki, and Youcef Tebbal. The complexity of dominating set reconfiguration. *Theor. Comput. Sci.*, 651(C):37–49, October 2016.

[6] Amer E. Mouawad, Naomi Nishimura, Venkatesh Raman, Narges Simjour, and Akira Suzuki. On the parameterized complexity of reconfiguration problems. *Algorithmica*, 78(1):274–297, May 2017.

[7] Daniel Lokshtanov, Amer E. Mouawad, Fahad Panolan, M.S. Ramanujan, and Saket Saurabh. Reconfiguration on sparse graphs. *Journal of Computer and System Sciences*, 95:122 – 131, 2018.

Modeling the neural network responsible for song learning

Silvia PAGLIARINI[1,2,3]
Directeur : Xavier HINAUT[1,2,3]**, Arthur LEBLOIS**[3]

1. INRIA Bordeaux Sud-Ouest, Talence, France
2. Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux
3. Institut des Maladies Neurodéegéenéerative, CNRS UMR 5293, Université de Bordeaux, France.

Humans, like songbirds, learn by imitation from an early age (e.g. speech learning in children and song learning in songbirds): they are capable to reproduce an experienced sensory stimulus (e.g. a sound) by finding the appropriate motor command to reproduce it. My PhD project aims to build a bio-inspired model to describe imitative vocal learning in birds using either classical mathematical models (e.g. dynamical systems) or recent developments in artificial intelligence (e.g. Generative Adversarial Networks, the so-called GAN).

1 Introduction

During the first period of their life, babies and juvenile birds show comparable phases of vocal development: first, they listen to their parents/tutors in order to build a neural representation of the experienced auditory stimulus, then they start to reproduce the sound trying to make it closer and closer to their parents one. This phase of learning is called sensorimotor phase and is characterized by the presence of babbling, in babies, and *subsong*, in birds. It ends when the song *crystallizes*, i.e. when it becomes similar to the one produced by the adults [1, 7]. Moreover, it is possible to find analogies between brain pathways responsible for sensorimotor learning in humans and in bird: a vocal production pathway involves direct projections from auditory areas to motor neurons, and a vocal learning pathway is responsible for imitation and plasticity [2]. The behavioral studies and the neuroanatomical structure of the vocal control circuit in human and bird provides the basis for a bio-inspired model of vocal learning. In particular, birds have brain circuits exclusively dedicated to song learning, making them an ideal model for exploring the representation of vocal learning by imitation.

2 Vocal learning model

Taking into account biological evidences and computational limitations, many previous studies have attempted to implement imitative learning in computational models. As shown in Fig. 1, the representation of a minimal vocal learning model includes three feature spaces and the functional connections between them. The motor space contains the motor parameters (related to the anatomical structure of the vocal organ, or related to the sound structure). The sensory space contains the auditory stimulus (the real sound), which is generated by the motor control function. The goal space is the space of the sensory stimuli that the bird wants to reproduce. Two types of model can be defined, depending on how the goal space is defined. In the sensorimotor model with an action-perception loop (top panel in Fig. 1) the sensory stimuli are encoded via the sensory response function in the perceptual space (a low dimensional representation of the sound). This model potentially includes an inverse and a forward model: these models provide a bi-directional link between the perceptual space and the motor space. Alternatively, in the non-perceptual sensorimotor model (bottom panel in Fig. 1) there is a non-perceptual representation of the goals (so-called internal representation). A goal-to-motor model learns the connection from the internal representation to the motor space. In this scenario, and depending on the learning framework, the sensory response function could provide a

reward or an evaluation of the learning (for this reason there is an arrow starting from the sensory space, but without a specific output space) [9].

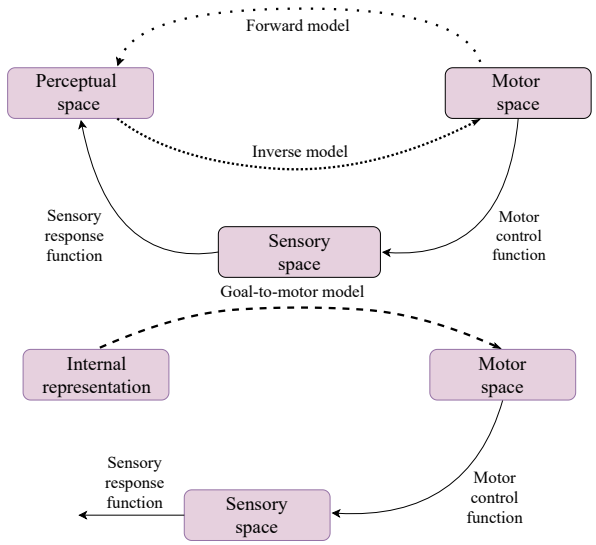

Figure 1: ***Top****: Sensorimotor model with a action-perception loop.* ***Bottom****: Non-perceptual sensorimotor model.*

2.1 Learning architecture

We started from a theoretical inverse model [6] where neurons are implemented as rate coding neurons and learning is driven by a normalized Hebbian learning rule, I introduced a non-linearity in the sensory response function to have stronger biologically hypothesis [5]. Moeover, I modified the normalization of the learning rule to achieve convergence of learning [8].

2.2 Motor control

In order to have a complete model (i.e. able to perceive and produce), we aim to have a motor control function capable of reproducing sounds similar to real data (i.e. recordings of canaries and zebra finches, adults and juveniles). I'm pursuing two approaches to build this function. On the one hand, I define the sound as the solution of a dynamical system based on the syrinx (the laryngeal organ in birds) structure and on the features of the sound (e.g. frequency, pitch) [4]. On the other hand, I generate the sound using a generative adversarial network (GAN), WaveGAN [3]. Starting from a dataset of real data (song recordings from zebra finches and canaries), this model allows to obtain a low-dimensional representation of the sound, called latent space. To test the capability of the motor control function to reproduce a sound similar to the real ones, I analyze the features (e.g. duration, height, wiener entropy, maximum and minimum frequency) of the generated and real syllables.

3 Future directions

I aim to use the dynamical system and the WaveGAN described in Subsection 2.2 as a motor function to build a new sensorimotor model with action-perception loop, where the connections between perceptual space and motor space are learned via an inverse model. As a general perspective, the schema proposed in Fig. 1 could be applied not only to build and analyze models for a vocal task, but more generally for sensorimotor tasks (e.g. imitation of a gesture).

Acknowledgment

This work was supported by the Inria CORDI-S PhD fellowship grant.

References

[1] MS. Brainard, AJ. Doupe, *What songbirds teach us about learning*, Nature, 2002.

[2] M. Chakraborty and ED. Jarvis, *Brain evolution by brain pathway duplication*, Philosophical Transactions of the Royal Society B: Biological Sciences, 2015, The Royal Society.

[3] C. Donahue et al., emphAdversarial audio synthesis, arXiv preperint, 2018.

[4] K. Doya, TJ. Sejnowski *A Computational Model of Avian Song Learning*, The new cognitive neurosciences (2nd ed.) Gazzaniga, M. S. (Ed.), 2013, Cambridge, MA, US: The MIT Press.

[5] R. Hahnloser, A. Kotowicz *Auditory representations and memory in birdsong learning*, Current opinion in neurobiology, 2010, Elsevier.

[6] A. Hanuschkin et al., *A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models*, Frontiers in neural circuits, 2013, Frontiers.

[7] PK. Kuhl, *Early language acquisition: cracking the speech code*, Nature reviews neuroscience, 2004, Nature.

[8] S. Pagliarini et al, *A bio-inspired model towards vocal gesture learning in songbird*, ICDL-Epirob, 2018, IEEE.

[9] S. Pagliarini et al., *Vocal Imitation in Sensorimotor Learning Models: a Comparative Review*, Journal of transactions on cognitive and developmental sysyems, SI: continual unsupervised sensorimotor learning, 2019, in press.

6D Pose Estimation without Retraining

Giorgia PITTERI
Directeurs : Aurélie Bugeau, Vincent Lepetit

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Estimating the 6D pose (the 3D translation and 3D rotation) of objects is one of the fundamental problems of Computer Vision, with many applications such as augmented reality and robotics. Although there have been years of research on 3D pose estimation problem, it still remains unsolved especially in complex scenarios (symmetrical and untextured objects, occlusions, different lighting conditions). In this thesis, we first propose a method to predict the 6D pose of symmetrical objects, then we propose two different approaches to predict the 6D pose of new objects without any additional learning nor training images of them.

1 Introduction

3D object detection and pose estimation are of primary importance for tasks such as robotic manipulation, virtual and augmented reality. They have been the focus of intense research in recent years, mostly due to the advent of Deep Learning based approaches and the possibility of using large datasets for training such methods. Methods relying on depth data acquired by depth cameras are robust. Unfortunately, active depth sensors are power hungry or sometimes it is not possible to use them. Deep Learning (DL) provides powerful techniques to estimate the 6D pose of the object from simple color images but many challenges remain, such as : varying light conditions, presence of clutter and occlusions, ambiguous objects texture, shape, and symmetries. In particular, symmetrical objects are frequent in industrial contexts, as illustrated in Figure 1.

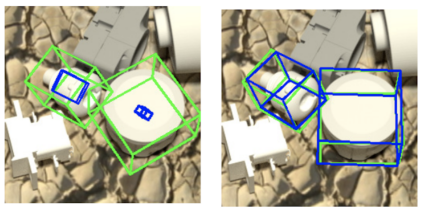

Figure 1: *Predicted poses for symmetrical objects with a naive DL approach (left) and our method (right).*

Furthermore, relying on supervised Machine Learning approaches means that for each new object, the networks have to be retrained on many different images of this object. Even if domain transfer methods allow for training such methods with synthetic images instead of real ones at least to some extent, such training sessions take time, and avoiding them is highly appealing.

In this thesis, we thus focus on two main problems that have not been (or almost not) addressed yet:

- how to deal with symmetrical objects;
- how to estimate the pose of new objects without any additional learning.

2 Handling Object Symmetries

As shown in Figure 2, when an object has some symmetry, there exist one or more rigid motions such that, if we apply these rigid motions to the object pose, the appearance of the object is preserved. We call such rigid motions 'equivalent rotations'. This means that two images of a symmetrical object can be identical without corresponding to the same pose. In this case, there is no function that maps an image to a unique pose. If a deep network was trained to predict the pose minimizing a cost function between the ground truth poses and the predicted poses, it would converge to a model predicting the average of the possible poses for

an input image, which is of course meaningless. In [2], we propose an efficient and simple solution that relies on the normalization of the pose rotation. We first map the equivalent rotations to a unique rotation, which we call a canonical rotation. This is sufficient for objects with a continuous set of symmetries such as cylinders and spheres. For objects with a discrete set of symmetries, this introduces some discontinuities around the symmetrical axes. We avoid these discontinuities by introducing a partition of the rotation space made of two subsets. For each subset, we train a different network to predict the pose. Our approach is general and can be used with any 6D pose estimation algorithm. It was presented at 3DV, the main conference on 3D Computer Vision, in 2019.

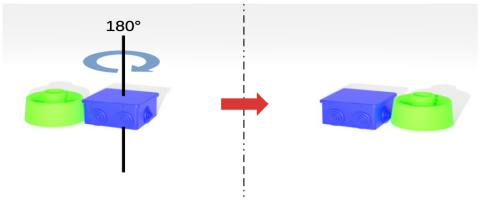

Figure 2: *An "equivalent rotation" shared by two objects. The appearance of the objects remain the same after this rotation.*

3 Generic 3D Corners for 6D Pose Estimation of New Objects without Retraining

In [1], we predict the 6D pose of new objects without additional learning nor training images for new objects. We consider a scenario where CAD models for the target objects exist, but not necessarily training images. This is often the case in industrial settings, where an object is built from its CAD model. These objects often share some similar parts (corners, edges, etc.). Our key idea is therefore to learn the 6D pose of the objects by predicting the pose of these parts. We decided to focus on corners since they are a dominant common pattern. We learn to detect them and estimate their 3D poses during an offline stage. We follow a deep learning approach and train an object detector framework (FasterRCNN) on a small set of objects to learn to detect corners and predict their 3D poses from color images. The 3D pose of a corner is predicted in the form of a set of 2D reprojections of 3D virtual points. Because of its symmetries, the 3D pose of a corner is often ambiguous, and defined only up a set of equivalent rotations. We therefore introduce a robust algorithm that estimates the object's 3D pose by matching its corners on the CAD model with their counterparts detected in the input image, considering the multiple possible 3D poses of the detected corners. We finally rely on a final check that exploits the full 3D geometry of the objects. The advantages of our approach make it particularly appealing for industrial contexts. This work was presented at a workshop at ICCV, one of the top-tier conferences in computer vision, in 2019.

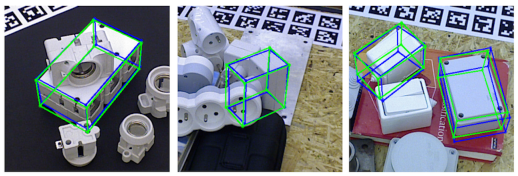

Figure 3: *Predicted poses for new objects without retraining.*

4 Beyond Corners: Dense Correspondences

The approach described in the previous section requires an expert to select parts on the new objects. To relax this requirement, and make our solution more general, we are developing a new approach based on dense correspondences between the input color image and the CAD model, instead of sparse corner correspondences. It is based on a rigid motion-invariant embedding of local 3D geometry, which allows us to generalize to different global shapes. For points at the surface of objects, our embedding can be computed from the CAD model, but it can also be predicted from images. This establishes dense correspondences between 3D points on the CAD model and 2D locations of new input images. From these correspondences, we can then robustly estimate a 3D pose.

References

[1] G. Pitteri, S. Ilic, V. Lepetit, *CorNet: Generic 3D Corners for 6D Pose Estimation of New Objects without Retraining*, ICCVW, 2019.

[2] G. Pitteri, M. Ramamonjisoa, S. Ilic, V. Lepetit, *On Object Symmetries and 6D Pose Estimation from Images*, 3DV, 2019.

[3] T. Hodaň, P. Haluza, Š. Obdržálek, J. Matas, M. Lourakis, X. Zabulis, *T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects*, WACV, 2017.

Partial Order Reduction for Networks of Timed Automata

Govind RAJANBABU
Directeurs : Igor WALUKIEWICZ, Frédéric HERBRETEAU

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Directeur : B Srivathsan

Chennai Mathematical Institute

Cyber physical systems are usually modelled as networks of timed automata operating in parallel. A huge hurdle in the efficient verification of such systems is the combinatorial explosion of state space due to concurrency. We study this problem and propose a new partial order-reduction algorithm for some classes of networks of timed automata which can cover various practical modeling questions.

Cyber-physical systems are present all around us and various tasks in our life are aided and assisted by their involvement. Their use is extensive - from control units in aircrafts and powergrids, to pacemakers and anti-lock braking systems in automobiles. These systems usually have some *specifications*, expressed as constraints in the way they behave. Quite often, these systems are safety-critical - even a remote chance of erroneous behaviour would result in a huge loss to life and property. Therefore, it is very crucial that we ensure that such systems always satisfy their specifications. *Verification* is an area of reasearch dedicated to developing procedures to address this key issue.

Verification of cyber-physical systems involves modeling the components of the system as abstract mathematical objects, known as *timed automata*. A timed automaton [1] is a finite state automaton equipped with a set of non-negative real-valued variables called *clocks*, initially set to zero and increasing at the same rate. The transitions of the automaton are associated with a conjunction of clock constraints, referred to as *guards*, where each clause in the conjunction involves the comparison of a clock with a natural number. Operationally, this means that the transition can be taken only if the value of the clocks satisfy the guard associated to the transition. In addition, sets of clocks can be *reset* ($i.e$, their values can be set to 0) on taking a transition. The erroneous behaviour of the system is modelled as reaching a special “bad” state of the timed automaton. The *reachability problem* for a timed automaton asks if there is a run of the automaton from an initial state to any given state. Therefore, checking whether the system has an erroneous execution translates to checking the reachability of the bad state in the timed automaton that models the system. The standard approach for solving this reachability problem involves exploring a directed graph, known as the *zone graph* of the timed automaton, and checking if a node corresponding to a bad state is reachable in this graph.

The aforementioned challenge is exacerbated by the problem of scalability. Since complex cyber-physical systems are modelled as several timed automata operating in parallel, verification of such systems result in a combinatorial explosion of the search space. We consider the *concurrent timed semantics* for the networks of timed automata, where the concurrency between the components of the network implies that there are a lot of interleavings of the same set of transitions. Many of these interleavings are in fact equivalent, and hence, need not all be explored. The drastic increase in the size of the search-space due to these interleavings needs to be addressed if we are to effectively carry out verification of complex cyber-physical systems which involve thousands of components.

A widely used technique used to combat the combinatorial explosion of search space is *partial order reduction* [6]. In general, the various partial order re-

duction techniques work by identifying a small part of state space whose exploration is sufficient to verify the system. Application of partial order reduction to networks of finite state automata has been extensively studied, but their extension to timed systems has remained a challenge due to the implicit synchronization induced by time. The main motivation of our work is to develop partial order reduction techniques for timed systems so as to obtain efficient verification procedures for networks of timed automata.

An approach in this front was the introduction of the notion of *local time semantics* by Bengtsson *et al.* in [3], which is an alternate semantics for timed automata. They propose that the *local zone graph*, which is the zone graph obtained from local time semantics, has some nice properties which makes it better suited for partial order reduction than the classical global zone graph. Minea in his work [4] built on this idea and gave a partial order reduction algorithm for timed systems.

In our work, we revisit local time semantics for networks of timed automata and point out a flaw in the existing approach [4] in applying partial order reduction for timed systems. Furthermore, we provide a new algorithm [8] that builds the local zone graph and give a new technique for guaranteeing termination. We point out that this graph is smaller than the global zone graph and also subsumes the gains observed by [5].

Unfortunately, when we apply our technique to make the local zone graph finite, we lose the "good" properties of the local zone graph which makes it suitable for partial order reduction. Consequently, we cannot apply partial order reduction directly on our finite local zone graph. To overcome this problem, we propose a better technique to ensure the exploration of termination of the local zone graph that preserves its good properties. We introduce the notion of *spread* of a network of timed automata and give some sufficient conditions to say when a network has bounded spread. We identify two classes of networks of timed automata, which we refer to as *global/local sytems* and *client/server systems*, and propose a partial order reduction algorithm for such networks of bounded spread. We also provide an evaluation of a prototype on some examples using the tool Tchecker [7].

References

[1] Rajeev Alur and David Dill. *A theory of timed automata*. Theoretical Computer Science, 126(2):183–235, 1994.

[2] Gerd Behrmann, Patricia Bouyer, Kim G. Larsen, and Radek Pelánek. *Lower and upper bounds in zone-based abstractions of timed automata*. International Journal on Software Tools for Technology Transfer, 8(3):204–215, 2006.

[3] Johan Bengtsson, Bengt Jonsson, Johan Lilius, and Wang Yi. *Partial order reductions for timed systems*. In CONCUR, volume 1466 of Lecture Notes in Computer Science, pages 485–500, 1998.

[4] Marius Minea. *Partial order reduction for model checking of timed automata*. In CONCUR, volume 1664 of Lecture Notes in Computer Science, pages 431–446. Springer, 1999.

[5] Ramzi Ben Salah, Marius Bozga, and Oded Maler. *On interleaving in timed automata*. In CONCUR, volume 4137 of Lecture Notes in Computer Science, pages 465–476. Springer, 2006.

[6] Parosh Aziz Abdulla, Stavros Aronis, Bengt Jonsson, and Konstantinos Sagonas. *Source Sets: A Foundation for Optimal Dynamic Partial Order Reduction*. In Journal of the ACM, volume 64, 2017.

[7] Frederic Herbreteau and Gerald Point. TChecker. https://github.com/tictac/tchecker.

[8] R Govind, Frédéric Herbreteau , B Srivathsan, and Igor Walukiewicz. *Revisiting local time semantics for networks of timed automata*. In CONCUR 2019.

Temporal graph properties and mobility-related problems

Jason SCHOETERS
Advisor: Arnaud CASTEIGTS

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

During my PhD, we studied temporal graph properties and mobility-related problems. In dynamic graphs, multiple useful temporal properties can be defined, such as connectivity over time. Naturally, different problems can be defined around these properties, such as how to preserve some property through edge deletion, i.e., how to obtain a temporal spanner. Another problem, joining mobility to dynamic graphs, is how to ensure in a swarm of autonomous entities, some temporal property on the underlying dynamic graph, through the use of some distributed mobility model. Concerning mobility, one can refine well-known combinatorial problems to better represent physical constraints, and possibly obtain a distinct interesting problem. For example, adding some acceleration constraints to the Traveling Salesman Problem can be powerful enough to result in different optimal results.

1 Temporal graph properties

A dynamic graph (also called temporal graph or time-varying graph) is defined as a graph in which edges can appear and disappear over some defined timeline. If one supposes the evolution of the graph known over the entire timeline, one can define some temporal properties in these graphs. Casteigts et al. [3] defined an extensive hierarchy of useful temporal graph properties (see Figure 1). These properties are necessary or sufficient for some algorithms to work efficiently.

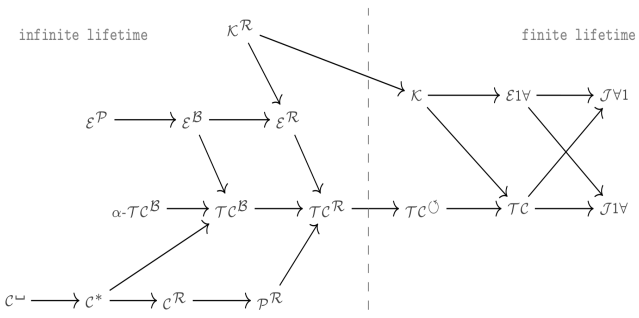

Figure 1: *Updated hierarchy of temporal properties [3].*

1.1 Temporal spanners

One of the more important temporal properties is *temporal connectivity*. Kempe *et al.* [5] asked whether, when given a temporally connected graph, one can always obtain a sparse subset of edges that preserves this temporal connectivity, *i.e.*, if one can always obtain a sparse temporal spanner. Negative results on this question have been obtained on graphs in general [5], and on dense graphs [1]. We showed the first positive result: for any complete dynamic graph, on can always obtain a sparse $O(n \log n)$-size spanner [4] (see Figure 2).

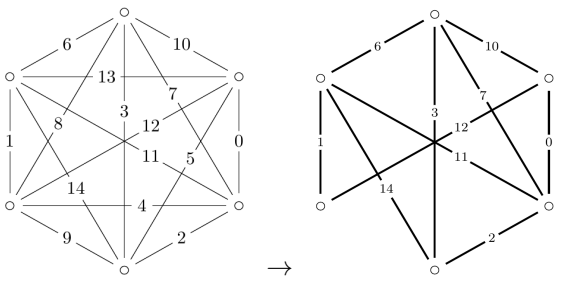

Figure 2: *A complete dynamic graph, and one of its temporal spanners.*

Future work includes narrowing down when one can, or cannot, find a sparse spanner in dynamic graphs, the limit lying somewhere between our positive result and

Axiotis *et al.*'s negative result [1]. Also, the question is still open if our result is tight, or whether for example $\Theta(n)$-size spanners can always be found in temporal cliques.

1.2 Properties through mobility

In the setting of Mobile Ad Hoc Networks (MANET or simply swarms), mobility of the swarm's entities play a significant role in the structure of the underlying communications graph, since the entities often have very short communication ranges (see Figure 3).

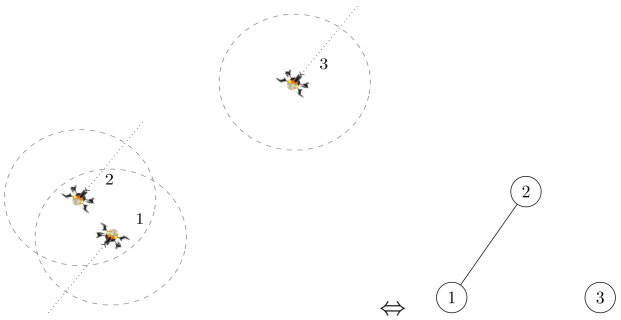

Figure 3: *Some mobile entities with limited communication range, and the corresponding communication graph. Note how this graph evolves over time, possibly losing edge* $(1, 2)$ *and gaining edge* $(1, 3)$ *in the future.*

We propose distributed mobility models which, when applied to the swarm's entities, generate an underlying dynamic graph with a desired temporal property. These properties can be necessary for some algorithm to function correctly in the swarm. For example, the three possible broadcasts in dynamic graphs, namely shortest, fastest, and foremost broadcast, are highly related to the three properties $\mathcal{E}^{\mathcal{R}}$, $\mathcal{E}^{\mathcal{B}}$ and $\mathcal{E}^{\mathcal{P}}$ [3]. Future work includes verifying whether standard MANET missions (such as surveillance or exploration) are delayed, or even obstructed by a temporal property's mobility models. Note these problems are also mobility-related problems (Section 2).

2 Mobility-related problems

Due to the popularity of research revolving around the use of MANET and drones (see Subsection 1.2), older problems can be updated *w.r.t.* mobile autonomous entities which might be used to accomplish various tasks (*e.g.* shortest path problems and pathfinding algorithms). The study of the effects of mobility itself, such as speed, acceleration and/or inertia forces, can even result in new and interesting problems in their own right, as opposed to being a simple adaption.

2.1 Vector Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is the problem of visiting a set of cities in minimum time. The standard problem does not take into account acceleration however, which can be observed for example by TSP allowing sharp angle turns between cities. Dubins TSP adds some useful constraints [6], but no acceleration. We added simple acceleration constraints on the visiting entity, known as Racetrack acceleration constraints [2] (see Figure 4).

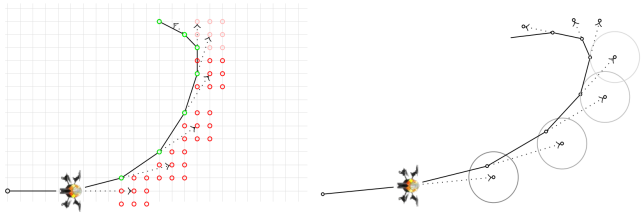

Figure 4: *Example trajectory of an entity with Racetrack acceleration constraints, and our continous version.*

These acceleration constraints modelize speed, acceleration, and inertia forces in a simple way in the TSP. The resulting problem is distincly different from TSP however, as the added acceleration constraints are strong enough to possibly result in a different optimal visit order than the corresponding TSP without acceleration. We call this the Vector Traveling Salesman Problem, and show NP-completeness for the problem, as well as several insights and algorithmic results.

References

[1] K. Axiotis, D. Fotakis, *On the size and the approximability of minimum temporally connected subgraphs*. arXiv preprint arXiv:1602.06411. 2016.

[2] M.A. Bekos, T. Bruckdorfer, H. Förster, M. Kaufmann, S. Poschenrieder, T. Stüber, *Algorithms and insights for racetrack*. Theoretical Computer Science. 2018 Nov 14;748:2-16.

[3] A. Casteigts, *A Journey through Dynamic Networks (with Excursions)*. Habilitation thesis, Université de Bordeaux, June 4, 2018.

[4] A. Casteigts, J.G. Peters, J. Schoeters, *Temporal cliques admit sparse spanners*. 46th Int. Colloquium on Automata, Languages, and Programming (ICALP), 2019.

[5] D. Kempe, J. Kleinberg, A. Kumar, *Connectivity and inference problems for temporal networks*. Journal of Computer and System Sciences. 2002 Jun 1;64(4):820-42.

[6] K. Savla, E. Frazzoli, F. Bullo, *On the point-to-point and traveling salesperson problems for Dubins' vehicle*. InProceedings of the 2005, American Control Conference, 2005. 2005 Jun 8 (pp. 786-791). IEEE.

Emergence of working memory in recurrent neural networks

Anthony Strock
Supervisors : Nicolas Rougier and Xavier Hinaut

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux
INRIA Bordeaux Sud-Ouest, Bordeaux, France
Institut des Maladies Neurodégénératives, Université de Bordeaux, CNRS UMR 5293, Université de Bordeaux

In this thesis project we propose to explore potential neural mechanisms for working memory. Working memory is what enables us to maintain information for a short time in order to help us solving tasks such as mental calculations. In our approach we consider a randomly connected population of neurons, a.k.a. a reservoir, that learns to perform tasks requiring working memory. Later on, we "dissect" these reservoirs to explain what are the mechanisms implementing working memory that emerged from learning. Our study focuses on two concrete cases: (1) gating, a generic mechanism for maintaining information, and (2) the mental calculations we learn in primary school.

One way to solve tasks is to learn by heart all the input-output associations. For instance, that is precisely what kids do in primary schools when they learn how to sum or multiply all the single digit numbers. In this way, they are able to learn a few hundred arithmetic facts. However, given the amount of information that would need to be memorized, it's not a solution that can be used for everything. For instance, in order to know how to sum or multiply double digit numbers, a kid would have to remember a few ten thousand arithmetic facts, i.e. 100 times more, and 100 times more for each additional digit. Instead we rely on a procedural mechanism to break down the task into easier tasks and sequentialise the resolution. For instance the sum of double digit numbers can be broken down using arithmetic facts on single digit numbers. One way to do this in practice would be to first add the two-unit digits, and then add the tenth digits with the potential carry over. That is precisely what most of kids learn in school by using a paper and pencil as memory. Similarly, before knowing by heart all the arithmetic facts with single-digit numbers, they would begin by first using their fingers as memory to count. What is interesting in the context of my thesis is that later on they become able to do all this "in their heads", i.e. without any external support for memory. To do all this, they use what is called working memory. Working memory is what allows them to maintain information (e.g. intermediate results) for a short period of time and to process it (e.g. sum two single digit numbers). The general aim of my thesis is to explore the potential neural mechanisms supporting this working memory.

To intuit such mechanisms one could start by exploring the huge literature of behavioral study of working memory. In practice there is a plethora of known effects on working memory and it is difficult to know where to start, i.e. which effects should be incorporated into our mechanism, although some people have tried to rank the importance of such effects [1]. Moreover, it's still not entirely clear which areas of the brain play a role in this mechanism. Prefrontal Cortex (PFC) seems to be central, because (1) damage to PFC leads to working memory deficits [2], and (2) recordings in the PFC have exhibited working memory representations[2]. But the role of PFC is not yet completely understood and a lot of other regions of the brain including cortical and subcortical regions seem to play a role [3]. Beyond that, how the PFC would maintain information is not yet fully understood either. What we propose in this thesis is to put aside all these a priori in order to explore at a rather algorithmic level the mechanisms that could support working memory. For this, we use a very simple class of Recurrent Neural Networks (RNN), a.k.a. reservoirs, which appears to have properties similar to those of PFC [4]. Our general approach consists in teaching these reservoirs how to solve tasks

that require working memory, and then dissect what mechanism this network takes to implement working memory.

The first part of my thesis focused on gating, a generic mechanism for maintaining information. The main idea of gating is to consider that the mechanism which maintains information can be divided into two sub-mechanisms: (1) a mechanism that acts as a gate, opens when new information needs to be stored and closes otherwise, and (2) a mechanism which maintain (resp. update) information when the door is closed (resp. open). It has been hypothesized that the two mechanisms could be carried out by two different regions of the brain [5]. The Basal Ganglia (BG) would perform the first while the PFC would perform the second. In this first part we focused on the second mechanism that could be performed by the PFC. One hypothesis in the literature is that in order to maintain information, neurons do it through sustained neural activity [6]. Alternatively, the information could be encoded in a rather dynamic way, i.e. the representation of the information held in the neurons changes over time [7]. Both hypotheses seem to be supported by empirical evidences, and in [8] we have precisely shown how we can build a robust RNN which can support both hypotheses.

The second part of my thesis was aimed at integrating the former with another model of the literature for long-term memories [9]. The aim was twofold: (1) to explore a potential mechanism to transfer between short and long-term memories, and (2) to explore a mechanism of transfer between memories encoded in the form of neural activity and memories encoded in the form of synaptic weights. In fact, it has been found empirically that information maintained on a temporary basis was not only encoded through neural activity, but could also be through properties of the synapses (e.g. synaptic weights) [7]. In addition, the model for long-term memories we considered have been introduced with a concept of operations. These operations allow a memory to be combined with other memories to form new ones. Therefore in [10] we proposed a mechanism of transfer between memories and we have described what are the effects of the combinations of long-term/synaptic memories on the short-term/neural-activity memory.

In the last part of my thesis we went back to the basic example we presented earlier, namely mental calculations. The idea here was to explore how procedural mechanisms such as the algorithms we learn in school to add numbers, interact with working memory.

References

[1] K. Oberauer, S. Lewandowsky, E. Awh, G. D. Brown, A. Conway, N. Cowan, ... G. Ward, *Benchmarks for models of short-term and working memory.*, Psychological Bulletin, (2018).

[2] S. Funahashi, *Working Memory in the Prefrontal Cortex*, Brain Sciences, (2017)

[3] T. B. Christophel, C. P. Klink, B. Spitzer, P. R. Roelfsema, J. D. Haynes *The Distributed Nature of Working Memory*, Trends in Cognitive Sciences, (2017)

[4] P. Enel, E. Procyk, R. Quilodran, P. F. Dominey, J. X. O'Reilly, *Reservoir Computing Properties of Neural Dynamics in Prefrontal Cortex*, PLOS Computational Biology, (2016)

[5] R. C. O'Reilly, M. J. Frank, *Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia*, Neural computation, (2006)

[6] M. L. Leavitt, D. Mendoza-Halliday, J. C. Martinez-Trujillo, *Sustained Activity Encoding Working Memories: Not Fully Distributed*, Trends in Neuroscience, (2017)

[7] M. G. Stokes, *'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework*, Trends in cognitive sciences, (2015)

[8] A. Strock, X. Hinaut, N. P. Rougier, *A Robust Model of Gated Working Memory*, Neural Computation, (2020)

[9] H. Jaeger, *Using Conceptors to Manage Neural Long-Term Memories for Temporal Patterns*, The Journal of Machine Learning Research, (2017)

[10] A. Strock, X. Hinaut, N. P. Rougier, *Transfer between long-term and short-term memory using Conceptors*, submitted, (2020)

2 – Mathématiques

Mathematical analysis for an non-local evolutionary-epidemic system arising in population dynamics

Lara ABI-RIZK
Directeur : Jean Baptiste BURIE et Arnaud DUCROT

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

In this work we study the travelling wave solutions for a spatially distributed system of equations modelling the evolutionary epidemiology of plant-pathogen interaction. Here the mutation process is described using a non-local convolution operator in the phenotype space. Using dynamical system ideas coupled with refined estimates on the asymptotic behaviour of the profiles, we prove that the wave solutions have a rather simple structure. This analysis allows us to reduce the infinite dimensional travelling wave profile system of equations to a fourth dimensional ode system. The latter is used to prove the existence of travelling wave solutions for any wave speed larger than a minimal wave speed $c_\star$, provided some parameters condition expressed using the principle eigenvalue of some integral operator. It is also used to prove that any travelling wave solution connects two determined stationary states.

1 Evolutionary epidemics

In this summary we present a general epidemiological model for plant diseases, formulated to study the evolution of phenotypic traits of plant pathogens in response to host resistance.
More precisely, the model reads as follows

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} = \Lambda - u(t,x) - u(t,x)\displaystyle\int_{\mathbb{R}^M} \beta(z) w(t,x,z)dz, \\ \dfrac{\partial v(t,x,y)}{\partial t} = u(t,x)w(t,x,y) - \mu_v v(t,x,y), \\ \left(1 - \dfrac{\partial^2}{\partial x^2}\right) w(t,x,y) = L[v(t,x,\cdot)](y), \end{cases} \tag{1}$$

where

$$L[v(t,x,\cdot)](y) = \Theta(y)\int_{\mathbb{R}^M} J(y-y')\Theta(y')v(t,x,y')dy', \tag{2}$$

posed for time $t \in \mathbb{R}$, spatial position $x \in \mathbb{R}$ and phenotypic trait value $y \in \mathbb{R}^M$, with some given integer $M \geq 1$. The functions $u = u(t,x)$, $v = v(t,x,y)$ and $w = w(t,x,y)$ denote the densities of the healthy plant population, the infected plant population (i.e. the plant surface bearing spore colonies), and the population of spores produced in the environment, respectively.
Here, $\Lambda > 0$ is the influx of new healthy population density, $\mu_v > 0$ is the infected plant death rates. The healthy population becomes contaminated by contact with the spores falling on the foliar surface that create a fungal colony with the transmission rate β, the parameter $\delta > 0$ is the spores deposition rate and $\Theta(y) = \sqrt{r(y)\beta(y)}$ represents the fitness function.The spores are produced by the colonies on the infected plant population of trait value y' with rate $r(y')$. The spores population mutates from trait y' to trait y proportionally to the kernel value $J(y-y')$.

2 Epidemic threshold and equilibria

Under some assumptions, the operator $L : \varphi \mapsto \Theta J * (\Theta\varphi)$ is irreducible, positive and compact in $L^p(\Omega)$ (and self-adjoint in L^2) so that spectral decomposition $(\varphi_n, \lambda_n)_{n\geq 1}$ with

$$\lambda_0 > \lambda_1 \geq \cdots \geq \lambda_n \to 0 \text{ as } n \to \infty \text{ and } \varphi_1 > 0.$$

The existence of the endemic equilibrium, and consequently of a travelling wave connecting the disease free equilibrium (DFE) to the endemic one, is related to the epidemic threshold $\mathcal{R}_0$ given by

$$\mathcal{R}_0 = \lambda_1 \Lambda/\mu_v$$

.

(i) Single equilibrium when $\mathcal{R}_0 \leq 1$, the DFE ($U = \Lambda$);
(ii) Two equilibria when $\mathcal{R}_0 > 1$, the DFE and an endemic one,

$$(U,V,W) = (U^*, V^*\varphi_1(x), W^*\varphi_1(x)),$$

with

$$(U^*, V^*, W^*) = \left(\frac{\Lambda}{\mathcal{R}_0}, \frac{\mathcal{R}_0 - 1}{\lambda_1 \beta_1}, \frac{\mathcal{R}_0 - 1}{\beta_1} \right), \qquad (3)$$

$$\beta_1 = \langle \beta, \varphi_1 \rangle_{L^2} .$$

3 Travelling wave

Definition:
An entire solution $(u(t,x), v(t,x,y), w(t,x,y))$ of (1) is said to be a travelling wave with speed $c > 0$ if it has the following form

$$(u(t,x), v(t,x,y), w(t,x,y)) \equiv (U(\xi), V(\xi,y), W(\xi,y)),$$

$$\xi = x + ct,$$

and if (U, V, W) satisfies :

- $(U(\xi), V(\xi,y), W(\xi,y))$ positive and bounded.
- the profile (U, V, W) satisfies

$$\lim_{\xi \to -\infty} \begin{pmatrix} U(\xi) \\ V(\xi,y) \\ W(\xi,y) \end{pmatrix} = \begin{pmatrix} \Lambda \\ 0 \\ 0 \end{pmatrix}$$

in $\mathbb{R} \times L^1(\mathbb{R}^M) \times L^1(\mathbb{R}^M)$.

4 Main Mathematical results

4.1 Non existence

We start by expressing conditions on the epidemic threshold $\mathcal{R}_0$ and on the speed c for which no travelling wave solution exists. Our result reads as follows

(i) If $\mathcal{R}_0 \leq 1$ then Problem (1) does not has any travelling wave solution.

(ii) If $\mathcal{R}_0 > 1$ then Problem (1) does not admit any travelling wave solution with wave speed $c \in (0, c_\star)$,

$$c_\star > 0, \quad \text{minimal wave speed.}$$

4.2 Qualitative properties

This result is concerned with some qualitative properties of the travelling wave solutions for (2), when they exist, in the remaining case $\mathcal{R}_0 > 1$ with wave speed $c \geq c_\star$.

Assume that $\mathcal{R}_0 > 1$. Let (U, V, W) be a TW profile for some speed $c \geq c_\star$, then there exist two smooth real functions $\hat{V} = \hat{V}(\xi) > 0$ and $\hat{W} = \hat{W}(\xi) > 0$ such that

(i) $(V(\xi,y), W(\xi,y)) \equiv \left(\hat{V}(\xi)\varphi_1(y), \hat{W}(\xi)\varphi_1(y) \right)$.

(ii) The function $\left(U(\xi), \hat{V}(\xi), \hat{W}(\xi) \right)$ satisfies the following system of equations

$$\begin{cases} c\dfrac{d}{d\xi} U(\xi) = \Lambda - U(\xi) - \beta_1 U(\xi) \hat{W}(\xi), \\ c\dfrac{d}{d\xi} \hat{V}(\xi) = U(\xi)\hat{W}(\xi) - \mu_v \hat{V}(\xi), \qquad \xi \in \mathbb{R}, \\ \lambda_1 \hat{V}(\xi) + \left(\dfrac{d^2}{d\xi^2} - 1 \right) \hat{W}(\xi) = 0, \end{cases} \qquad (4)$$

and the limit behaviors at $\xi = \pm\infty$:

$$\lim_{\xi \to -\infty} \begin{pmatrix} U(\xi) \\ \hat{V}(\xi) \\ \hat{W}(\xi) \end{pmatrix} = \begin{pmatrix} \Lambda \\ 0 \\ 0 \end{pmatrix},$$

$$\lim_{\xi \to \infty} \begin{pmatrix} U(\xi) \\ \hat{V}(\xi) \\ \hat{W}(\xi) \end{pmatrix} = \begin{pmatrix} U^* \\ V^* \\ W^* \end{pmatrix},$$

wherein the positive steady state (U^*, V^*, W^*) is defined in (3). With $\lambda_1 = \rho(L)$ and we have set $\beta_1 = \int_{\mathbb{R}^M} \beta(y)\varphi_1(y)dy > 0$.

The above result shows that the wave profiles of (1) have a simple shape and that the investigation of such special solutions reduces to the study of the 4-dim ODE problem (4) with the expected limiting behavior, $\mathcal{R}_0 > 1$ and for wave speed $c \geq c_\star$.
The existence of solutions for this reduced system of equations is ensured by our next result.

4.3 Existence of travelling waves for $\mathbf{c \geq c_\star}$

Assume that $\mathcal{R}_0 > 1$. Then for each wave speed $c \geq c_\star$, Problem (4) admits – at least – a positive solution

References

[1] Abi Rizk, L. J.-B.Burie, A. Ducrot (2019). Travelling wave solutions for a non-local evolutionary-epidemicsystem. Journal of Differential Equations Volume 267, Issue 2, 5 July 2019, Pages 1467-1509. https://doi.org/10.1016/j.jde.2019.02.012

Computing the Hilbert class field of a family of primitive quartic CM fields

Jared ASUNCION
Directeur : Andreas Enge, Marco Streng

Université de Bordeaux / Universiteit Leiden

1 Introduction

An abelian extension of a number field K is a Galois extension L/K whose Galois group $\mathrm{Gal}(L/K)$ is abelian. The theorem of Kronecker-Weber states that every finite degree abelian extension L of $\mathbb{Q}$ is contained in $\mathbb{Q}(\exp(2\pi i\tau) : \tau \in \mathbb{Q})$, the field obtained by adjoining to $\mathbb{Q}$ the values of the analytic map $\tau \mapsto \exp(2\pi i\tau)$ evaluated at $\mathbb{Q}$. The twelfth out of the twenty-three problems Hilbert posed in 1900, roughly speaking, asks whether a statement analogous to Kronecker-Weber can be made for number fields K different from $\mathbb{Q}$.

1.1 CM theory

One important tool in attacking this problem is the theory of complex multiplication (CM), studied by Shimura [5, 6] during the second half of the 20th century. CM theory works for a specific family of number fields, called CM fields. A CM field K is a totally imaginary number field, which is a quadratic extension of a totally real number field K_0.

1.2 Class field theory

For a CM field K and any positive integer n, class field theory proves the existence of a family of fields denoted by $H_K(n)$, indexed by positive integers n, such that a finite degree abelian extension L of K, is contained in $H_K(f)$ for some positive integer f. This field $H_K(n)$ can be defined more precisely for general number fields.

1.3 The field generated by CM

CM theory provides a way to explicitly construct an abelian extension $\mathrm{CM}_K(n)$ of a CM field K that has the following properties:

- $\mathrm{CM}_K(n) \subseteq H_K(n)$
- $\mathrm{CM}_K(n)$ can be obtained by adjoining special values of analytic functions, in a similar fashion to the Kronecker Weber theorem.

The simplest example of a CM field K is an imaginary quadratic number field. In this case, one can prove that $\mathrm{CM}_K(n)$ is exactly $H_K(n)$ thereby solving Hilbert's twelfth problem for the case of imaginary quadratic number fields. However, the statement $\mathrm{CM}_K(n) = H_K(n)$ does not hold for a general CM field.

2 The Hilbert class field as a subfield of the field generated by CM

Let F be a primitive quartic CM field. The 'primitive' condition is equivalent to saying that F is either not Galois or cyclic Galois. Shimura [4, Theorem 2] proved that

$$H_F(1) \subseteq \Xi_F(m) := H_{F_0}(m)\mathrm{CM}_F(m) \qquad (\tilde{\star})$$

for some positive integer m. A more precise reformulation of $\tilde{\star}$ is available in the upcoming preprint, along with a more precise treatment of aspects of CM theory.

2.1 An $\mathfrak{m}$ for which $\tilde{\star}$ holds

Let F be a primitive quartic CM field. Let $\mathcal{O}_F$ be its ring of integers. Let

$$\mathrm{Cl}_F(1) = \frac{\{\text{ideals of } \mathcal{O}_F\}}{\{\text{principal ideals of } \mathcal{O}_F\}}$$

Let S be a finite set of prime ideals of K such that S contains all prime ideals of K above 2 and the group $\mathrm{Cl}_K(1)/\langle S\rangle$ has odd order. Let

$$\mathcal{P} = \{p : p \text{ is a rational prime below } \mathfrak{p} \text{ for some } \mathfrak{p} \in S\}$$

Let P be the product of all elements of $\mathcal{P}$. We prove that if $m = 4P$ then $\tilde{\star}$ holds. Our result thereby makes the aforementioned theorem of Shimura more explicit.

2.2 Given $\mathfrak{m}$, does $\tilde{\star}$ hold?

The above value for m is usually too large for computational purposes. It is not surprising to find a smaller integer for which $\tilde{\star}$ is true. This begs the question: given an integer m, does $\tilde{\star}$ hold? If $F \subseteq F' \subseteq H_F(m)$, denote $\mathrm{Gal}(H_F(m)/F')$ by $G(F')$. Using Galois correspondence we find that this question is equivalent to asking whether

$$G(H_F(1)) \supseteq G(H_{F_0}(m)) \cap G(\mathrm{CM}_F(m)) \qquad (\tilde{\star}\tilde{\star})$$

holds. Computing $G(H_F(1))$ and $G(H_{F_0}(m))$ have been well-studied parts of class field theory. We give an algorithm to compute the generators and the group structure of $G(\mathrm{CM}_F(m))$. Algorithms to compute intersections of subgroups exist [1], thereby enabling us to determine whether $\tilde{\star}\tilde{\star}$ holds, given an integer m.

3 Computing the Hilbert class field

Using the theory we have established, we find an algorithm to find $H_F(1)$ given that $\tilde{\star}$ holds for $m = 1$ or $m = 2$. The algorithm involves computing theta constants, explicit Shimura reciprocity [8] and a computation of $\mathrm{CM}_F(m)$ for $m \in \{1, 2\}$ [9, 3, 2].

In the practical side of things, we have also implemented our algorithm. Our implementation uses PARI/GP [10] to do the class group and unit group computations, which computes $G(\mathrm{CM}_F(m))$. Moreover, it uses RECIP SAGE package [7, 8] which finds defining polynomials of the primitive elements $\alpha_1, \alpha_2, \alpha_3$ such that $H_F(1) = F(\alpha_1, \alpha_2, \alpha_3)$ by computing theta constants and utilising explicit Shimura reciprocity.

The algorithm given by Kummer theory, which finds $H_F(1)$ for any number field F, is implemented in PARI/GP as `bnrclassfield`. When the degree $[H_F(1) : F]$ is large, our implementation of the CM algorithm has been tested to perform significantly faster than the current `bnrclassfield` implementation.

References

[1] Henri Cohen. *Advanced topics in computational number theory*. Graduate Texts in Mathematics. Springer, New York, 2000.

[2] Craig Costello, Alyson Deines-Schartz, Kristin Lauter, and Tonghai Yang. Constructing abelian surfaces for cryptography via rosenhain invariants. *LMS Journal of Computation and Mathematics*, 17(A):157–180, 2014.

[3] Andreas Enge and Emmanuel Thomé. Computing class polynomials for abelian surfaces. *Experimental Mathematics*, 23:129–145, 2014.

[4] Goro Shimura. On the class-fields obtained by complex multiplication of abelian varieties. *Osaka Math. J.*, 14(1):33–44, 1962.

[5] Goro Shimura. *Introduction to the Arithmetic Theory of Automorphic Functions*. Kanô memorial lectures. Princeton University Press, 1971.

[6] Goro Shimura. *Abelian Varieties with Complex Multiplication and Modular Functions*. Princeton Mathematical Series. Princeton University Press, 2016.

[7] W. A. Stein et al. *Sage Mathematics Software (Version 9.0)*. The Sage Development Team, 2020. `http://www.sagemath.org`.

[8] Marco Streng. An explicit version of shimura's reciprocity law for siegel modular functions. 2011.

[9] Marco Streng. *Complex multiplication of abelian surfaces*. PhD thesis, 2010.

[10] The PARI Group, Univ. Bordeaux. *PARI/GP version 2.11.2*, 2019. Available from `http://pari.math.u-bordeaux.fr/`.

Étude de la répartition des automorphismes de Frobenius

Alexandre Bailleul
Directeur : Florent JOUVE

Institut de Mathématiques de Bordeaux, CNRS UMR 5251, Université de Bordeaux

On s'intéresse à la répartition des nombres premiers, et plus généralement des idéaux premiers de corps de nombres satisfaisant certaines propriétés arithmétiques. Ces propriétés sont encodées dans des automorphismes dits de Frobenius. Des asymétries inattendues apparaissent lorsque l'on cherche à comparer la fréquence à laquelle l'une ou l'autre de ces propriétés se produit.

1 Répartition des nombres premiers

L'étude de la répartition des nombres premiers parmi les entiers naturels est un problème remontant à l'Antiquité (Euclide avait montré l'existence d'une infinité de nombres premiers) et qui continue à susciter beaucoup de recherche de nos jours. Le théorème fondamental concernant cette répartition est le Théorème des Nombres Premiers (Hadamard, de la Vallée-Poussin, 1896) :

$$\pi(x) := |\{p \leq x\}| \underset{x\to+\infty}{\sim} \mathrm{Li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t}.$$

Une question plus fine que l'on peut se poser à propos des nombres premiers est celle de leur répartition dans les progressions arithmétiques. Existe-t-il une infinité de nombres premiers dont le chiffre des unités est 7 (c'est-à-dire appartenir à la progression arithmétique $10\mathbb{N}+7$) ? Un célèbre théorème de Dirichlet (1837) répond par l'affirmative à cette question. Que se passe-t-il pour d'autres chiffres des unités ? Hormis pour 2 et 5, le chiffre des unités d'un nombre premier doit se trouver parmi $1, 3, 7$ ou 9. Il n'y a a priori aucune raison que l'un de ces chiffres apparaisse plus souvent que les autres. À la suite du théorème des nombres premiers, de la Vallée-Poussin a montré en 1899 la généralisation suivante : pour tout entier $q \geq 2$, et tout entier a premier avec q, on a

$$\pi(x; q, a) := |\{p \leq x \mid p \equiv a \bmod q\}| \underset{x\to+\infty}{\sim} \frac{\pi(x)}{\varphi(q)},$$

où $\varphi(q)$ est le nombre d'entiers entre 1 et q premiers avec q. Autrement dit, si a et b sont premiers avec q, il y a (asymptotiquement) autant de chance qu'un nombre premier soit congru à a modulo q ou à b modulo q. Dans le cas de la congruence modulo 10, on obtient qu'un nombre premier a autant de chance d'avoir un chiffre des unités 7 que 9 par exemple.

2 Biais de Tchebychev et courses de nombres premiers

Malgré ce résultat d'équirépartition, qui prédit une sorte d'harmonie dans la répartition des nombres premiers dans différentes classes de congruence, il existe parfois une asymétrie entre certaines de ces classes. C'est ce qu'a observé Tchebychev en 1853, entre les classes $1 \bmod 4$ et $3 \bmod 4$. Tchebychev prédit que $\pi(x; 4, 3)$ est "la plupart du temps" supérieur à $\pi(x; 4, 1)$ (cf. Figure 2).

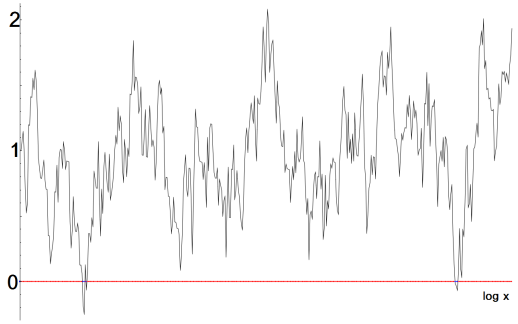

Figure 1: $\frac{\pi(x;4,3)-\pi(x;4,1)}{\sqrt{x}\log x}$, $10^4 \leq x \leq 10^8$ *(Daniel Fiorilli)*

Ce phénomène, appelé biais de Tchebychev, a finalement pu être quantifié et démontré (sous certaines conjectures naturelles) en 1994 par Rubinstein et Sarnak dans [5] : l'inégalité $\pi(x;4,3) > \pi(x;4,1)$ se produit "$\approx 99,59\%$ du temps". La démonstration utilise des fonctions holomorphes, appelées fonctions L de Dirichlet, qui permettent d'appréhender, à travers la répartition de leurs zéros, la répartition des nombres premiers dans les progressions arithmétiques. L'explication mise en lumière par l'analyse de Rubinstein et Sarnak est que ce biais se produit car 1 est un carré modulo 4 tandis que 3 ne l'est pas.

3 Généralisations

Si L/K est une extension galoisienne de corps de nombres (c'est-à-dire d'extensions finies de $\mathbb{Q}$), on peut associer à chaque idéal premier $\mathfrak{p}$ de K (sauf un nombre fini) un élément $\mathrm{Frob}_\mathfrak{p}$ de $\mathrm{Gal}(L/K)$, le groupe de Galois de L/K, appelé automorphisme de Frobenius de cet idéal. Celui-ci est obtenu en remontant dans $\mathrm{Gal}(L/K)$ l'automorphisme de Frobenius de l'extension résiduelle de corps finis $(\mathcal{O}_L/\mathfrak{P})\,/\,(\mathcal{O}_K/\mathfrak{p})$, où $\mathfrak{P}$ est n'importe quel idéal premier de L divisant $\mathfrak{p}$. Au lieu de comparer des nombres premiers en progressions arithmétiques, on peut comparer les idéaux premiers de K qui admettent différents automorphismes de Frobenius. Par exemple dans l'extension $\mathbb{Q}\left(e^{\frac{2i\pi}{q}}\right)/\mathbb{Q}$, l'automorphisme de Frobenius associé au nombre premier p est entièrement déterminé par la classe de congruence de p modulo q, et on retrouve le problème des courses de nombres premiers modulo q. La transposition des résultats de Rubinstein et Sarnak a été faite par Ng dans [4]. En général, des phénomènes nouveaux se produisent, liés à l'existence de zéros en $1/2$ de fonctions L plus générales que celles de Dirichlet, des fonctions L d'Artin. L'objet de mon article [1] est de mettre en évidence l'influence de ces zéros sur le biais de Tchebychev entre idéaux premiers. J'ai notamment montré l'existence de deux suites d'extensions de $\mathbb{Q}$ dont les groupes de Galois sont des groupes de quaternions généralisés $\mathbb{H}_{2^n}$ qui exhibent des biais de Tchebychev dans des directions opposés, à cause de l'existence pour l'une et de l'absence pour l'autre de zéros en $1/2$ de fonctions L d'Artin.

Si L/K est une extension galoisienne de corps de fonctions de courbes algébriques sur le corps fini $\mathbb{F}_q$, il existe une théorie proche de celle des corps de nombres permettant de se poser des questions similaires [2]. Mon travail actuel consiste à étudier les différences avec le contexte des corps de nombres (influence des genres des courbes, phénomènes de dépendance linéaire entre zéros de fonctions L) et adapter des résultats connus pour les corps de nombres à ce contexte.

Je travaille également à la simplification de l'étude des phénomènes de type biais de Tchebychev à travers le développement d'un langage probabiliste permettant d'appréhender les phénomènes oscillatoires sous-jacents.

References

[1] A. Bailleul, *Chebyshev's bias in dihedral and generalized quaternion Galois groups*. Prépublication, https://arxiv.org/abs/2001.06671, 2020.

[2] B. Cha, B.-H. Im, *Chebyshev's bias in Galois extensions of global function fields*, J. Number Theory **131** (2011), 1875–1886.

[3] H. Davenport, *Multiplicative Number Theory*, Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, 74. Springer-Verlag, New York, 2000. xiv+177 pp.

[4] N. Ng, *Limiting Distributions and Zeros of Artin L-functions*, Thèse de doctorat, 2004.

[5] M. Rubinstein, S. Sarnak, *Chebyshev's bias*, Experiment. Math., **3** (1994) 173–197.

Auto-propulsion et interaction d'ailes battantes dans des écoulements visqueux

Luis BENETTI RAMOS[1,2,3]
Directeurs : Michel BERGMANN[1,2]**, Angelo IOLLO**[1,2]
Co-encadrant : Olivier MARQUET[3]

[1] *Université de Bordeaux, IMB, UMR 5251, F-33400 Talence, France*
[2] *Equipe-projet Memphis, Inria Bordeaux-Sud Ouest, F-33400 Talence, France*
[3] *ONERA, The French Aerospace Lab - F-92190 Meudon, France*

Dans cette thèse on se propose d'étudier numériquement l'auto-propulsion d'ailes battantes dans des écoulements visqueux, sous l'angle de la stabilité linéaire comme non-linéaire. À l'aide de la résolution des équations de Navier-Stokes fortement couplées au mouvement de la structure, notre but est de comprendre les mécanismes physiques responsables de (i) la génération de forces propulsives et d'un mouvement d'avancement cohérent d'une aile seule, ainsi que (ii) les états d'interaction hydrodynamiques observés dans une allée infinie d'ailes battantes.

1 Contexte

La propulsion par battement d'ailes, utilisée par exemple par les oiseaux et les poissons à travers le mouvement de leurs ailes et nageoires, est une grande source d'inspiration pour la conception innovante de véhicules maritimes, amphibies et micro-drones. Ce mécanisme d'auto-propulsion résulte des forces de frottement visqueux et de l'accélération du fluide environnant engendrées par le couplage entre le mouvement de l'aile et l'écoulement fluide. Ce dernier dépend énormément de la gamme de Reynolds considérée. Dans le cas des micro-drones, l'écoulement fluide engendré par l'auto-propulsion de l'aile battante se situe dans une gamme intermédiaire de nombre de Reynolds (1-1000) [1], où les non-linéarités et l'inertie du fluide sont importantes. Des études récentes montrent que ces deux éléments sont essentiels pour la génération de forces propulsives, le mouvement d'avancement cohérent qui lui succède [2] et par l'augmentation de l'efficacité énergétique et la stabilité des formations des ailes battantes en propulsion collective [3]. Une bonne compréhension de ces phénomènes physiques est essentielle pour atteindre les impératifs de performance dans la conception de mécanismes basés sur la propulsion par battement.

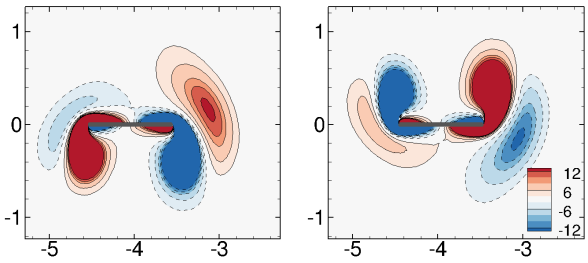

Figure 1: *Contours de vorticité: Auto-propulsion d'une aile battante dans un fluide visqueux initiellement au repos.*

2 Démarche, déroulement et principaux résultats obtenus

On étudie la dynamique linéaire et non-linéaire du système couplé fluide/aile en auto-propulsion, forcé par un mouvement de battement périodique de grande amplitude imposé à l'aile, dans le cas de deux problèmes modèles: la locomotion horizontale d'une aile

dans un fluide au repos et la propulsion d'une allée infinie d'ailes.

Dans le premier cas, on s'est intéressé à observer le comportement en auto-propulsion horizontale d'une l'aile symétrique à mouvement vertical imposé. Via la résolution numérique des équations de Navier-Stokes couplées à la dynamique horizontale de l'aile, trois différents régimes non-linéaires ont été observés : des solutions (i) non-propulsives, (ii) propulsives en avancement unidirectionnel et (iii) associées à un mouvement de va-et-vient périodique, ces dernières représentées dans la Figure 1. Pour comprendre la génération de forces et l'apparition de ces différentes solutions, nous avons étudié la stabilité au sens de Floquet des solutions périodiques non-propulsives. L'analyse de Floquet permet, d'une part, de caractériser via les multiplicateurs de Floquet, la stabilité de cet état de base, et d'autre part caractériser les mécanismes physiques responsables des comportements non-linéaires observés via les modes propres. Nos résultats de stabilité ont correctement prédit les seuils d'apparition des régimes propulsifs unidirectionnels ainsi que des mouvements de va-et-vient. On a montré que ces solutions sont liées à l'apparition des modes instables de Floquet, respectivement à pulsation nulle et à une basse fréquence. De la même manière, les modes propres obtenus nous ont permis d'identifier les mécanismes physiques responsables de la génération de forces propulsives et du déclenchement du mouvement de va-et-vient pour certaines fréquences.

Le deuxième problème abordé concerne la propulsion collective d'une allée infinie d'ailes battantes dans un canal. Dans ce cas, la configuration étudiée est confinée par des parois solides en haut et en bas. Pour contourner ce problème, des méthodes de domaine fictif [5] ont été implémentées pour prendre en compte le mouvement de la structure. Des simulations non-linéaires temporelles ont été menées et deux paramètres qui contrôlent l'interaction entre les ailes ont été explorés: la fréquence de battement et l'écartement entre les ailes. Des résultats obtenus pour différents écartements à une fréquence fixe sont représentés dans la Figure 2. On observe d'abord que pour un écartement très grand entre les ailes les effets d'interaction ne sont plus présents (Figure 2 b), et la vitesse moyenne du système tend vers la vitesse d'une aile seule. Pour des écartements inférieurs, la vitesse moyenne de l'allée peut devenir inférieure ou supérieure à celle d'une aile seule. Pour certains écartements, on peut trouver deux solutions stables différentes: l'une à vitesse de propulsion supérieure et l'autre inférieure à une aile seule. La stabilité de ces solutions est liée à l'interaction entre l'allée de tourbillons générée par les ailes précédentes et les tourbillons éjectés au bord d'attaque ou de fuite de l'aile considérée (Figure 2 (a)). Pour aller plus loin dans la compréhension des mécanismes hydrodynamiques responsables de la stabilité des solutions obtenues, on essaye d'accéder à des solutions non-linéaires instables, via une méthode d'équilibrage harmonique (Time Spectral Method, TSM [4]). Ces branches de solutions instables supposées seront suivies par une méthode de continuation.

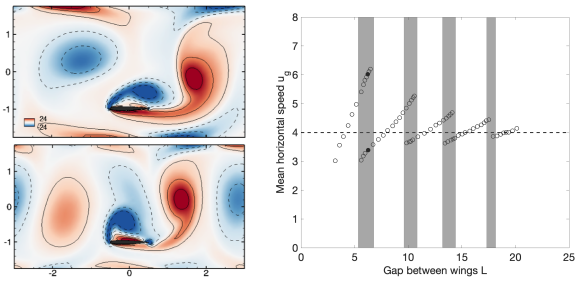

Figure 2: *Contours de vorticité (gauche) et vitesse moyenne horizontale (droite) pour la propulsion d'une allée infinie d'ailes battantes. En (b) la vitesse moyenne d'une aile seule est représentée par une ligne traitillée*

3 Perspectives

La dernière année de la thèse sera consacrée à l'approfondissement des deux problèmes étudiés. Dans le premier cas, une analyse approfondie des mécanismes physiques responsables des modes instables basée sur le concept d'endogénéité de l'opérateur couplé fluide/solide linéarisé est en cours. Dans le cas de l'allée infinie d'ailes battantes, on s'intéressera à l'optimisation de la vitesse de propulsion, soit par une stratégie non-linéaire consistant à calculer des perturbations optimales soit par une stratégie d'optimisation linéaire originale consisterait à regarder les valeurs singulières de l'opérateur TSM linéarisé.

References

[1] D. Faux *et al EPL* **121**: 66001, 2018.

[2] S. Alben & M. Shelley, *PNAS*, **102**: (32), 11163–11166, 2005.

[3] Becker, A. Masoud, H. Newbolt, J. Shelley, M. Ristroph, L., *Nature Commnications* **6**: 1-8, 2015.

[4] Mundis, N. Mavriplis, D. *Journal of Computational Physics* **345**, 2017.

[5] Boffi, D., Hecht, F., Pironneau, O., *Numerical Methods for PDEs: State of the Art Techniques* 129-145, 2018.

Learning of Pathological Radar Clutter using Supervised and Unsupervised Classification Methods

Yann CABANES
Directeurs : Marc ARNAUDON & Jérémie BIGOT, Co-encadrant : Frédéric BARBARESCO

Institute of Mathematics of Bordeaux, CNRS, Bordeaux INP, University of Bordeaux, Thales LAS France, DGA AID

We propose a method to classify radar clutter from radar data using unsupervised classification algorithms. Thus new radars will be able to use the experience of other radars, wich will improve their performances: learning pathological radar clutter can be used to fix some false alarm rate created by strong echoes coming from hail, rain, waves, mantains, cities; it will also improve the detectability of slow moving targets, like drones, which can be hidden in the clutter, flying close to the landform.

1 Introduction

Radar installation on a new geographical site is long and costly. We would like to shorten the time of deployment by recognizing automatically pathological clutters with past known diagnosed cases. This requirement will become more important because new generation radar sensors will be mobile and should work on the move and self-adapt to the environment. The objective is therefore to develop machine learning algorithms to recognize specific clutter characteristics from their Doppler Spectrum.

To recognize pathological radar environments using a pulse-Doppler radar, we need to extract very precise Doppler information from a very small series of pulses (around 10). In this context, the classical FFT (Fast Fourier Transform) is not satisfactory due to its low resolution output for such small time series. To overcome this difficulty, we propose a clustering method based on the signals Toeplitz covariance matrices.

To begin with, we briefly introduce the radar data which we intend to analyze. For simplicity, we first consider one fixed direction in which a radar sends radio waves and we subdivide this direction into cells. The radar sends a burst of radio waves in a direction and then receives the echoes. For each echo we measure its amplitude r and phase ϕ, so that it can be represented by a complex number $u = re^{i\phi}$. As a result, the original radar observation value of each cell is a complex vector $\mathbf{u} = [u(0), u(1), ..., u(n-1)]^T$, where n is the number of radio waves emitted in each burst and $\cdot^T$ denotes the matrix transposition. We now try to extract sharp Doppler information from this short time series. Instead of using directly the original observation vector $\mathbf{u}$ of each cell, we assume it to be a realization of a centered stationary complex Gaussian process and identify it with its covariance matrix $\mathbf{R} = \mathbb{E}[\mathbf{u}\,\mathbf{u}^H]$, where $\cdot^H$ denotes the complex matrix conjugate transpose. In other words, the new observation value for each cell is a covariance matrix estimation, which is Toeplitz due to the assumption of stationarity of the process. Then our clustering problem can be summarized as follows: regroup in a same cluster the cells having close Toeplitz covariances matrices.

In order to achieve this clustering problem, three fundamental issues should be addressed. The first one is how to estimate a Toeplitz covariance matrix from each original radar observation vector $\mathbf{u}_i$. The second one is to define a metric on the set of Toeplitz

covariance matrices. The third one is to adapt our clustering algorithms to the newly defined metric space of Toeplitz covariance matrices.

2 The model

We assume that the signal can be modeled as a centered stationary autoregressive Gaussian process of order $n-1$:

$$u(k) + \sum_{i=1}^{n-1} a_i^{n-1}\, u(k-i) = w(k) \tag{5}$$

where a_i^{n-1} are the prediction coefficients and w is the prediction error.

3 Equivalent representations

According to [4] the Levinson algorithm gives us the following bijection:

$$\begin{aligned} &\mathcal{T}_n^+ \rightarrow \mathbb{R}_+^* \times \mathbb{D}^{n-1} \\ &R_n \mapsto (p_0, \mu_1, ..., \mu_{n-1}) \end{aligned} \tag{6}$$

where $\mathcal{T}_n^+$ denotes the set of Positive Definite Hermitian Toeplitz matrices of size n; $p_0 = r_0$ is the averaged quadratic power and $\mu_j = a_j^j$ $(1 \leq j \leq n-1)$ are the reflection coefficients, where a_i^j denotes the i^{th} coefficient of the Gaussian autoregressive model of order j. It is therefore equivalent to estimate the coefficients $(p_0, \mu_1, ..., \mu_{n-1})$ and R_n.

In practice, we use an algorithm called the Burg algorithm [6] to estimate the reflection coefficients.

4 The Kähler metric

We define the metric on the set $\mathcal{T}_n^+$ of Toeplitz Hermitian Positive Definite matrices as coming from the restriction of the Fisher metric on the manifold of complex Gaussian distributions with zero means to the manifold of complex Gaussian distributions with zero means, Toeplitz Hermitian Positive Definite covariance matrices and null relation matrices.

According to bijection (6), we can represent a Toeplitz Hermitian Positive Definite matrix T_i by the corresponding coefficients $(p_{0,i}, \mu_{1,i}, ..., \mu_{n-1,i})$. The following distance has been introduced by F. Barbaresco in [7] on the set $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ to make this bijection an isometry:

$$\begin{aligned} d^2_{\mathcal{T}_n^+}(T_1, T_2) = &\quad d^2_{\mathcal{T}_n^+}\left((p_{0,1}, \mu_{1,1}, ..., \mu_{n-1,1}), (p_{0,2}, \mu_{1,2}, ..., \mu_{n-1,2})\right) \\ = &\quad n \log^2\left(\frac{p_{0,2}}{p_{0,1}}\right) + \sum_{l=1}^{n-1} \frac{n-l}{4} \log^2\left(\frac{1 + \left|\frac{\mu_{l,1} - \mu_{l,2}}{1 - \mu_{l,1}\mu_{l,2}^*}\right|}{1 - \left|\frac{\mu_{l,1} - \mu_{l,2}}{1 - \mu_{l,1}\mu_{l,2}^*}\right|}\right) \end{aligned} \tag{7}$$

Note that the metric on the product space $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ is a product metric, which greatly simplifies the computations. The equations of the geodesics of the set $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ endowed with the Kähler metric are described in [5].

5 Clustering

The difficulty here is to adapt classical clustering algorithms to the Riemannian manifold described previously. Interested readers can find detailed information about our work in our articles [1, 2, 3].

References

[1] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Unsupervised Machine Learning for Pathological Radar Clutter Clustering: the P-Mean-Shift Algorithm", IEEE, C&ESAR 2019, Rennes, France, 19-21 November 2019

[2] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Non-Supervised High Resolution Doppler Machine Learning for Pathological Radar Clutter", IEEE, RADAR 2019, Toulon, France, 23-27 September 2019

[3] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Toeplitz Hermitian Positive Definite Matrix Machine Learning based on Fisher Metric", IEEE, GSI 2019, Toulouse, France, 27-29 August 2019

[4] B. Jeuris and R. Vandrebril, "The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks", 2016

[5] Marc Arnaudon, Frédéric Barbaresco, Le Yang, "Riemannian Medians and Means With Applications to Radar Signal Processing", IEEE journal, August 2013.

[6] F. Barbaresco, "Super resolution spectrum analysis regularization: Burg, Capon and AGO-antagonistic algorithms", in EUSIPCO-96, Trieste, Italy, 1996, pp. 2005-2008

[7] F. Barbaresco, "Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median". In Matrix Information Geometry; Bhatia, R., Nielsen, F., Eds.; Springer, 2012; pp. 199–256.

Fonctions de Littlewood-Paley-Stein pour les opérateurs de Schrödinger et le laplacien de Hodge-de Rham sur des variétés non-compactes

Thomas Cometx
Directeur : El Maati Ouhabaz

Institut de Mathématiques de Bordeaux, CNRS UMR 5251, Université de Bordeaux

Les fonctions de Littlewood-Paley-Stein associées à des semi-groupes sont des objets importants en analyse harmonique sur les variétés Riemaniennes. Nous étudions les fonctions associées à deux opérateurs particuliers : l'opérateur de Schrödinger et le laplacien de Hodge-de Rham sur les 1-formes. Nous nous intéressons à leur continuité sur les espaces L^p et sur les liens de celles-ci avec la transformée de Riesz sur les variétés.

1 Fonctions de Littlewood-Paley-Stein

Sur une variété Riemannienne M, l'opérateur de Laplace-Beltrami est défini par $\Delta = d^*d$. Ici d est la dérivée extérieure (où le gradient de la fonction) et $d*$ est son adjoint pour le produit scalaire sur $L^2(M)$, munissant M de la mesure induite par la distance Riemannienne. Soit V une fonction localement intégrable sur M, on définit aussi les opérateurs de Schrödinger sur les fonctions sur M par $Lf = \Delta f + Vf$.

Dans son livre [2], Stein introduit les fonctions de Littlewood-Paley-Stein pour Δ. Plus généralement, on les définit pour L. La fonctionnelle verticale de Littlewood-Paley-Stein est définie pour tout fonction f lisse à support compact par

$$H_L(f)(x) = \left(\int_0^\infty |de^{-tL}f|_x^2 dt\right)^{1/2}.$$

La fonctionnelle horizontale est définie par

$$h_L(f)(x) = \left(\int_0^\infty |\frac{\partial}{\partial t}e^{-tL}f|^2 t dt\right)^{1/2}.$$

Les fonctionnelles h_L et H_L sont des applications sous-linéaires définies sur l'espace des fonctions C^∞ à support compact sur M. On se demande si elles se prolongent en des applications sous-linéaires continues sur les espaces $L^p(M)$.

On se pose la même questions pour d'autres fonctions de Littlewood-Paley-Stein, cette fois associées au laplacien de Hodge-de Rham sur les 1-formes. Ce laplacien est défini par $\overrightarrow{\Delta} = dd^* + d^*d$. Soit ω une 1-forme différentielle sur M, on définit comme précédemment les fonctions de Littlewood-Paley-Stein associées à $\overrightarrow{\Delta}$:

$$H_{\overrightarrow{\Delta}}(\omega)(x) = \left(\int_0^\infty |d^*e^{-t\overrightarrow{\Delta}}\omega|_x^2 dt\right)^{1/2},$$
$$h_{\overrightarrow{\Delta}}(\omega)(x) = \left(\int_0^\infty |\frac{\partial}{\partial t}e^{-t\overrightarrow{\Delta}}\omega|_x^2 t dt\right)^{1/2}.$$

Ici $|.|$ est la norme sur l'espace cotangent en x. Le laplacien de Hodge-de Rham sur les 1-formes peut être vu comme un opérateur de Schrödinger à potentiel vectoriel. En effet, $\overrightarrow{\Delta} = \nabla^*\nabla + R$ où ∇ est la connexion de Levi-Civita et R le tenseur de Ricci sur les 1-formes. Une des approches pour traiter le cas du laplacien de Hodge-de Rham est d'adapter les résultats que l'on obtient sur les opérateurs de Schrödinger en transposant les propriétés du potentiel V sur le tenseur de Ricci R.

2 Quelques résultats connus

Des intégrations par parties montrent que ces fonctions sont toujours bornées sur L^2. Par interpolation, elles sont donc bornées sur L^p pour p dans un certain intervalle non vide. Dans [2], Stein prouve que H_Δ est toujours borné sur L^p pour $p \in (1, 2]$. Il montre aussi que sur $\mathbb{R}^n$, H_Δ est borné pour $p \in (1, \infty)$.

Pour l'opérateur de Schrödinger V, beaucoup de résultats sont connus si $V > 0$: H_L est toujours borné pour $p \in (1, 2]$ et pour beaucoup de potentiels V on sait que H_L n'est pas borné dès que p est assez grand (par exemple, dans le cas de $\mathbb{R}^n$, dès que $p \geq n$).

3 Un exemple de motivation : les transformées de Riesz

Une des raisons pour lesquelles on étudie ces fonctions est d'obtenir la bornitude en norme L^p d'un autre opérateur : la transformée de Riesz. Celle-ci est définie par $d\Delta^{-1/2}$. Sa continuité sur L^p se traduit par l'existence de $C > 0$ tel que pour toute fonction C^∞ à support compact,

$$\int_M |df|_x^p dx \leq C \int_M |\Delta^{1/2} f|^p dx.$$

Par intégration par parties, l'égalité est toujours vraie pour $p = 2$. Les théorèmes d'interpolation classiques nous assurent alors que l'ensemble des valeurs de p pour lesquelles la transformée de Riesz est bornée est un intervalle non vide (éventuellement réduit à un singleton).

Dans [1], Coulhon et Duong montrent que si H_Δ est bornée sur L^p et si $h_{\overrightarrow{\Delta}}$ est bornée sur $L^{p'}$ (où $\frac{1}{p} + \frac{1}{p'} = 1$), alors la transformée de Riesz est bornée sur L^p. En étudiant l'adjoint de la transformée de Riesz, on montre de la même façon que dans [1] que si $H_{\overrightarrow{\Delta}}$ est bornée sur $L^{p'}$, alors la transformée de Riesz est bornée sur L^p (voir [3]).

Réciproquement, la continuité sur L^p de la transformée de Riesz implique la continuité de H_Δ. On peut aussi définir les transformées de Riesz associées aux opérateurs de Schrödinger par $dL^{-1/2}$ et alors la continuité L^p de $dL^{-1/2}$ implique celle de H_L.

4 Un résultat positif : le cas sous-critique

Une modification de la preuve de Stein dans [2] permet d'obtenir des résultats positifs concernant la continuité de H_L et $H_{\overrightarrow{\Delta}}$.

4.1 Le cas de l'opérateur de Schrödinger

On décompose V comme la différence de sa partie positive et de sa partie négative en l'écrivant $V = V^+ - V^-$. On dit que V^- est sous-critique s'il existe $\alpha \in (0, 1)$ tel que pour toute fonction f,

$$\int_M V^- f^2 dx \leq \alpha \int_M |df|_x^2 + V^+ f^2 dx.$$

Sous cette condition nous avons montré que H_L est borné sur L^p pour $p \in (p_0, 2]$ où $p_0 = \frac{2}{1+\sqrt{1-\alpha}}$.

4.2 Le cas du laplacien de Hodge-de Rham sur les 1-formes

Comme dans le cas de l'opérateur de Schrödinger, on décompose la courbure de Ricci comme différence de sa partie positive et négative en l'écrivant $R = R^+ - R^-$. On dit que R^- est sous critique s'il existe $\alpha \in (0, 1)$ tel que pour toute 1-forme ω,

$$\int_M < R^- \omega, \omega > dx \leq \int_M |\nabla \omega|_x^2 + < R^+ \omega, \omega >_x dx,$$

où $< ., . >_x$ est le produit scalaire sur l'espace cotangent en x. Sous cette condition et si le semi-groupe $e^{-t\overrightarrow{\Delta}}$ vérifie une inégalité maximale sur L^p, nous avons montré que $H_{\overrightarrow{\Delta}}$ est borné sur L^p pour $p \in (p_0, 2]$. Il s'en suit que la transformée de Riesz est bornée pour $p \in [2, p_0')$, ce qui donne un résultat positif quant à sa bornitude reposant sur peu d'hypothèses.

References

[1] T. Coulhon, X.T. Duong, *Riesz transforms for p > 2*, C. R. Acad. Sci. Paris, t. 332, Série I, p. 975–980, 2001.

[2] E.M. Stein, *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory*. Annals of Mathematics Studies, Princeton University Press, 1970.

[3] T. Cometx, *Littlewood-Paley-Stein functions for Hodge-de Rham and Schrödinger operators*. arXiv:1912.08471, 2019.

Coefficients de Fourier de formes modulaires de poids demi-entier

Corentin DARREYE
Directeurs : Florent JOUVE et Guillaume RICOTTA

Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, Talence, France

Depuis presque deux siècles, les formes modulaires apparaissent indépendamment dans différents contextes mathématiques et physiques. Leur étude est devenue un sujet à part entière qui constitue aujourd'hui un bagage d'outils analytiques particulièrement utiles en Théorie des Nombres. Pour ne citer qu'un exemple, les formes modulaires jouent un rôle crucial dans la preuve par Andrew Wiles du dernier théorème de Fermat. Lors de mon Doctorat, j'ai eu l'occasion d'étudier certains aspects des formes modulaires, notamment des questions d'équirépartition liées à leurs coefficients de Fourier.

1 Formes modulaires de poids entier

1.1 Quelques définitions

Une forme modulaire f est une série de Fourier définie sur le demi-plan de Poincaré de la forme

$$f(z) = \sum_{n\geq 0} a(n) e^{2i\pi z}$$

avec $z = x + iy$ et $y > 0$. De plus, elle doit vérifier une certaine périodicité par rapport à z que l'on peut résumer en la relation suivante

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z) \tag{8}$$

pour toute matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Ici, k est un entier fixé qu'on appelle le poids. Par ailleurs, on peut remplacer $\mathrm{SL}_2(\mathbb{Z})$ par n'importe lequel de ses sous-groupes dans la définition. On note alors $\mathcal{M}_k$ l'ensemble des séries (convergentes) qui vérifient (8) et $\mathcal{S}_k$ le sous-ensemble des formes pour lesquelles $a(0) = 0$.

La relation fonctionnelle (8) confère aux coefficients $a(n)$ des propriétés arithmétiques, ce qui donne une première justification de l'intérêt qu'on leur porte en Théorie des Nombres. Un exemple important est la série d'Eisenstein définie pour les entiers pairs $k \geq 4$ par

$$G_k(z) = \sum_{(c,d)\in\mathbb{Z}^2\setminus(0,0)} \frac{1}{(cz+d)^k} \in \mathcal{M}_k$$

et dont le $n^{\text{ième}}$ coefficient de Fourier est proportionnelle à $\sum_{d|n} d^{k-1}$ pour tout $n \geq 1$.

On peut montrer que $\mathcal{M}_k$ est un espace vectoriel de dimension finie et admet une unique base de la forme $\{G_k\} \cup \mathcal{B}_k$ où toute forme $f \in \mathcal{B}_k \subset \mathcal{S}_k$ (dite propre) a des coefficients qui vérifient

$$\sum_{n\geq 1} \frac{a(n)}{n^s} = \prod_p \left(1 - a(p)p^{-s} + p^{k-1-2s}\right)^{-1}$$

pour Re $s > \frac{k+1}{2}$. En particulier, les coefficients $a(n)$ sont multiplicatifs *i.e.* $a(mn) = a(m)a(n)$ pour m et n premiers entre eux.

1.2 Estimations sur les coefficients

Lorsqu'on travaille avec des formes modulaires, on est souvent amené à contrôler la taille de leurs coefficients ou de sommes de coefficients. Pour $f \in \mathcal{S}_k$, on a

$$\sum_{n\leq x} |a(n)|^2 \underset{x\to+\infty}{\sim} c_f x^k \tag{9}$$

où c_f est une constante explicite. Ponctuellement, on a $a(n) = O_\epsilon(n^{\frac{k-1}{2}+\epsilon})$ pour tout $\epsilon > 0$ d'après un résultat très profond de Deligne [2]. Peu de choses sont connues sur les minorations possibles de $|a(n)|$.

Une approche possible serait d'exploiter des estimées du type (9) mais où l'on somme sur un ensemble plus petit, par exemple une progression arithmétique. Récemment, de nombreux articles ont étudié la somme des $a(n)$ lorsque $n = a \pmod p$ et $n \leq x$. Le résultat le plus important en date est le suivant.

Théorème 1 ([3],[4]) Soit $f \in \mathcal{B}_k$. Pour $x > 0$, un nombre premier p et une classe $a \pmod p$, on pose

$$E(x,p,a) = \frac{1}{\sqrt{x/p}} \sum_{\substack{n \leq x \\ n = a \pmod p}} a(n) n^{-\frac{k-1}{2}}.$$

Si p et x tendent vers $+\infty$ avec $x^{1/2}\log x \leq p$ et $p = O_\epsilon(x^{1/2+\epsilon})$ pour tout $\epsilon > 0$, alors

$$\frac{1}{p-1}\sum_{a=1}^{p-1} E(x,p,a)^\nu \to \frac{\nu!}{2^{\nu/2}(\nu/2)!} c_f^{\nu/2} \qquad (10)$$

pour tout entier pair $\nu \geq 0$ (les moments impairs tendant vers 0). En conséquence, la variable aléatoire $E(x,p,\cdot)$ converge en loi vers $\mathcal{N}(0,c_f)$, la loi normale centrée de variance c_f .

2 Poids demi-entier

Imaginons maintenant que l'on puisse remplacer l'entier k dans (8) par $1/2$, $3/2$, $5/2$... Ceci est possible à condition d'être vigilant sur le choix d'une racine carrée de $cz+d$ et, modulo certains détails techniques, on peut alors définir les formes modulaires de poids demi-entier. De façon amusante, le premier exemple de forme modulaire à apparaitre dans la littérature est une forme de poids demi-entier. Il s'agit de la fonction thêta de Jacobi (1829), définie par

$$\theta(z) = \sum_{n\in\mathbb{Z}} e^{2i\pi n^2}$$

et qui est de poids $1/2$.

Ces formes sont par bien des points analogues à celles de poids entier mais leurs coefficients restent de manière générale plus mystérieux. Par exemple, l'estimation (9) est encore valable et donc les $a(n)$ sont en moyenne plus petit que $n^{\frac{k-1}{2}}$. Toutefois, on ne sait pas prouver que $a(n) = O_\epsilon(n^{\frac{k-1}{2}+\epsilon})$ pour tout $\epsilon > 0$ mais seulement $a(n) = O_\epsilon(n^{\frac{k-1}{2}+3/14+\epsilon})$ pour n sans facteur carré (fait dû à Iwaniec). De plus, bien qu'il existe une notion de forme propre similaire à celle en poids entier, les $a(n)$ ne sont en général plus multiplicatifs dans ce cas. Ce dernier point complique significativement l'étude des coefficients de formes de poids demi-entier.

Le projet principal de ma thèse a été d'établir le Théorème 1 dans le cas où f est de poids demi-entier. Cependant dans ces conditions, la méthode utilisée dans [3] ou [4] ne fonctionne plus et on trouve alors par différents moyens que l'estimation (10) sur les moments de $E(x,p,a)$ n'est plus vraie. On obtient en particulier le résultat suivant.

Théorème 2 ([1]) Si f est une forme propre de poids demi-entier, alors il existe une sous-suite de $E(x,p,\cdot)$ qui converge en loi vers la loi mixte $\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0,2c_f)$ avec δ_0 la masse de Dirac en zéro.

Les Théorèmes 1 et 2 résultent de propriétés d'équirépartition de certaines sommes d'exponentielles. Précisément, on réduit le problème à l'étude de

$$\mathrm{Kl}_p(a;\chi) = \frac{1}{\sqrt{p}}\sum_{u=1}^{p-1}\chi(u)e^{2i\pi(au+\bar{u})/p} \in [-2,2]$$

où $u\bar{u} = 1 \pmod p$ et $\chi(u) = 1$ ou $\left(\frac{u}{p}\right)$ suivant si on se place dans le cas du poids entier ou demi-entier. On montre alors que $\mathrm{Kl}_p(\cdot;\chi)$ tend en loi quand $p \to +\infty$ vers une loi de densité $\frac{1}{2\pi}\sqrt{4-t^2}dt$ si $\chi = 1$ et de densité $\frac{1}{2}\delta_0 + \frac{1}{2\pi}\frac{dt}{\sqrt{4-t^2}}$ sinon. C'est cette différence de comportement asymptotique qui est à l'origine de la différence entre les énoncés des deux Théorèmes.

Pour finir sur une petite application, l'étude des sommes de coefficients dans les progressions arithmétiques permet de donner des minorations sur $\pm a(n)$. On peut notamment montrer que pour tout $\epsilon > 0$ et pour tout $x > 0$ assez grand, il existe plus de $x^{4/7-\epsilon}$ progressions arithmétiques dans lesquelles au moins un coefficients $a(n)$ avec $n \leq x$ vérifie $a(n) > n^{-3/14-\epsilon}$ et de même avec $a(n) < -n^{-3/14-\epsilon}$. Ce résultat non-trivial met en évidence le caractère oscillatoire des coefficients de Fourier d'une forme modulaire, principe à la base de nombreuses propriétés arithmétiques.

References

[1] C. Darreye, *Fourier coefficients of modular forms of half-integral weight in arithmetic progressions*, available on arXiv, (2019)

[2] P. Deligne, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math., **43**,(1974) 273–307.

[3] É. Fouvry, S. Ganguly, E. Kowalski, Ph. Michel, *Gaussian distribution for the divisor function and Hecke eigenvalues in arithmetic progressions*, Comment. Math. Helv., **89**, (2014) 979–1014

[4] S. Lester, N. Yesha, *On the distribution of the divisor function and Hecke eigenvalues*, Israel J. Math., **212**, (2016) 443–472

Méthodes d'agrégation et désagrégation de programmes linéaires en nombres entiers

Gaël Guillot
Directeur : François CLAUTIAUX

Institut de Mathématiques de Bordeaux, A33, 351 Cours de la Libération, 33400 Talence, Inria

Dans cette thèse, nous nous intéressons à des problèmes d'optimisation combinatoire, et plus particulièrement aux problèmes pouvant se formuler sous la forme d'un programme dynamique et de contraintes linéaires additionnelles. Ces problèmes posent des difficultés considérables aux méthodes classiques d'optimisation. Les approches les plus prometteuses sont basées sur la relaxation de l'espace d'états. L'objectif de la thèse est d'unifier ces méthodes et de les améliorer grâce à des méthodes itératives d'agrégation et de désagrégation. Nous avons validé nos approches sur des problèmes de sac à dos, d'ordonnancement et de planification du traitement phytosanitaire de la vigne.

1 Contexte scientifique

Dans de nombreux domaines d'application de l'optimisation combinatoire (transports, planification, découpe, ordonnancement...), la structure du problème repose sur la consommation/production de ressources limitées comme le temps ou des matières premières, rendant possible leur modélisation via le paradigme de programmation dynamique [1]. Ces problèmes peuvent alors être représentés sous forme de la recherche de chemins dans un graphe dont chaque nœud correspond à un état du système (ou contrainte), et chaque arc à une transition liée à une décision (ou variable).

La formulation de grande qualité obtenue se fait au prix d'un très grand nombre d'états des sous-systèmes, qui est généralement exponentiel en fonction du nombre de ressources, et pseudo-polynomial en la consommation de ces ressources, ce qui interdit l'utilisation directe de ces reformulations en pratique. Pour résoudre ces problèmes, des méthodes basées sur des relaxations et des énumérations sont utilisées.

Plusieurs types de relaxations peuvent être utilisées. On peut relâcher complètement les contraintes additionnelles, mais on peut aussi utiliser des relaxations comme la relaxation Lagrangienne. La notion de relaxation de l'espace d'états a été introduite en 1979 [2]. L'espace d'états initial est projeté dans un espace d'états plus petit, ce qui permet d'obtenir un programme dynamique relâché avec un nombre d'états et de transitions moins importants. L'agrégation permet d'obtenir une borne mais le problème est résolu par une méthode classique d'énumération (branch-and-bound, algorithmes de labels).

A partir de ces méthodes, des méthodes itératives ont été mises en place pour converger vers l'optimum. Ibaraki introduit l'algorithme Successsive Sublimation Dynamic Program (SSDP) [3] qui utilise la relaxation Lagrangienne à chaque itération pour supprimer des transitions et de résoudre les relaxations avec des algorithmes de graphes. Cette méthode a été utilisé sur divers problèmes d'ordonnancement [4]. Une autre méthode itérative d'agrégation, Decremental State Space Relaxation (DSSR) [6], utilise le même processus itératif mais résout les relaxations par des algorithmes de labels. Ces deux méthodes permettent d'obtenir des résultats intéressant sur plusieurs problèmes, en ordonnancement et en routing.

2 Contribution

Un des objectifs de cette thèse est de proposer un formalisme générique pour ces méthodes de manière à

pouvoir généraliser les méthodes de la littérature. Nous avons identifié une structure commune aux différentes méthodes, et mis en avant leurs ingrédients principaux.

Pour permettre à ces méthodes d'être compétitives, plusieurs problématiques doivent être résolues: quelle relaxation choisir, quel espace d'état, comment raffiner la relaxation à chaque itération, quelle méthode de résolution choisir pour les relaxations, ...

Pour valider notre approche, nous avons appliqué l'une d'elles sur le problème de sac-à-dos temporel, un problème de sac-à-dos classique auquel on ajoute des dates d'entrées et de sorties à chaque objet. La méthode SSDP a montré des résultats compétitifs avec la littérature [5], notamment grâce à un choix de dimension basé sur l'évolution du réseaux, une énumération partielle des décisions possibles à chaque états et différents tests de dominances et de faisabilités.

Il existe quelques travaux dans la littérature sur ces méthodes, notamment ceux de [4] sur des problèmes d'ordonnancement. L'un des objectifs est de comparer les performances de notre implémentation générique avec une implémentation spécialisée sur ces problèmes, et d'améliorer les résultats grâce à l'hybridation de méthodes.

Enfin, ces méthodes ont été confrontées à des problèmes industriels moins classiques, par exemple un problème d'utilisation de traitements phytosanitaires sur des vignes. En utilisant une méthode de décomposition, les sous-problèmes peuvent être formulés comme des programmes dynamiques de grande taille.

References

[1] R. Bellman, *The theory of dynamic programming*. RAND Corp Santa Monica CA (No. RAND-P-550), 1954.

[2] N. Christofides, A. Mingozzi, P. Toth, *State-space Relaxations for Combinatorial Problems*. Imperial College Internal Report IC, OR, 79, 09, 1979.

[3] T. Ibaraki, *Successive sublimation methods for dynamic programming computation*. Annals of Operations Research, 11(1), 397-439, 1987.

[4] S. Tanaka, S. Fujikuma, M. Araki, *An exact algorithm for single-machine scheduling without machine idle time*. Journal of Scheduling, 12(6), 575-593, 2009.

[5] A. Caprara, F. Furini, E. Malaguti, E. Traversi, *Solving the temporal knapsack problem via recursive dantzig–wolfe reformulation*. Information Processing Letters, 116(5), 379-386, 2016.

[6] G. Righini, M. Salani, *New dynamic programming algorithms for the resource constrained elementary shortest path problem*. Networks: An International Journal , 51(3), 155-170, 2008.

Calcul Stochastique dans les variétés et applications aux inégalités fonctionnelles

Baptiste HUGUET
Directeurs : Marc ARNAUDON & Michel BONNEFONT

Institut de Mathématiques de Bordeaux, CNRS UMR 5251, Université de Bordeaux

Dans mes travaux de thèse, j'étudie des semi-martingales à valeurs dans des variétés riemanniennes. La compréhension de ces processus a de nombreuses applications en analyse. J'ai travaillé en particulier sur deux de ses applications : d'une part, l'obtention d'inégalités fonctionnelles de type Poincaré via l'étude des semi-groupes associés; d'autre part, l'étude de l'équation de Navier-Stokes via le problème de Brödinger.

1 Inégalité de Poincaré

1.1 Diffusions et semi-groupes

Les semi-martingales forment une large classe de processus stochastique. L'exemple le plus simple de cette famille est le mouvement brownien. Bien que ces processus soient assez irréguliers ($1/2^-$-hölderien), il est possible de construire un calcul différentiel pour lequel ils se comporte exactement comme les processus $\mathcal{C}^1$: c'est le calcul de Stratonovitch. Les diffusions sont des semi-martingales particulières auxquelles est associer un opérateur différentiel - appelé générateur - qui les caractérise. Le mouvement brownien est la diffusion de générateur $1/2\Delta$. Je travaille avec avec des diffusions de générateur $L = \Delta - \nabla V.\nabla$ où V est un potentiel lisse. À toute diffusion, on peut associé un semi-groupe $\mathbf{P}_t$, défini par la formule

$$\mathbf{P}_t f(x) = \mathbb{E}\left[f(X_t(x))\right],$$

où $X(x)$ est une diffusion de générateur L issue de x. Cet objet est bien défini car une diffusion est unique en loi. Pour toute fonction f, $\mathbf{P}_t f$ est la solution de l'équation $\partial_t u = Lu$ avec condition initiale f. Au-dessus d'une diffusion, on peut construire le transport parallèle déformé, noté W_t. C'est un isomorphisme entre $T_{X_0}M$ et $T_{X_t}M$ qui vérifie l'équation suivante :

$$\begin{cases} D_t W_t v = -\mathcal{M}^* W_t v dt \\ W_0 = id_{T_{X_0}M} \end{cases}, \tag{11}$$

où D_t est la dérivée covariante le long de la diffusion et $\mathcal{M}^* = Ric^{\#} + \nabla^2 V$. Le transport parallèle déformé est aussi une diffusion mais à valeurs dans l'espace tangent TM. Son générateur sur les 1-formes, noté L^W, vérifie une importante propriété de commutation : $dL = L^W d$ où d désigne la différentielle. En tant que diffusion, W engendre aussi un semi-groupe sur les 1-formes que l'on notera $\mathbf{Q}_t$:

$$\langle \mathbf{Q}_t \alpha, v\rangle = \mathbb{E}\left[\langle \alpha, W_t v\rangle\right].$$

Une troisième manière de comprendre le transport parallèle déformé est démontré dans [2] : c'est la dérivée spatiale d'un flot "convenable" de la diffusion X.

1.2 Entrelacement et inégalité de Poincaré

On considère une probabilité μ qui a une densité de la forme $exp(-V)$. On recherche des condition sur le potentiel pour avoir l'inégalité suivante - dite inégalité de Poincaré :

$$Var_\mu(f) \leq c \int_M |df|\, d\mu.$$

Quel est le lien entre cette inégalité et les objets précédemment définis? Un calcul assez simple montre que l'on a :

$$Var_\mu(f) = \int_0^{+\infty} \int_M \langle df, d\mathbf{P}_t f\rangle \, d\mu \, dt.$$

La compréhension de $d\mathbf{P}_t f$ est donc importante. La relation de commutation entre les générateurs et le lien exposé dans [2] suggèrent une relation d'entrelacement entre les semi-groupes :

$$d\mathbf{P}_t f = \mathbf{Q}_t df.$$

Sous l'hypothèse classique de courbure-dimension, on peut montrer que l'entrelacement a lieu et on obtient le résultat souhaité. En effet, cette hypothèse est équivalente à une minoration du potentiel $\mathcal{M}$, ce qui entraine

une borne sur $|W_t|_2$ via le lemme de Grönwall. Mon travail a été de contourner cette hypothèse. La méthode utilisée dans [1], pour le cas euclidien, consiste à tordre le transport W et étudier le nouveau semi-groupe associé. Cela permet d'obtenir une relation d'entrelacement et parfois une inégalité de type Poincaré sans pour autant vérifier le critère de courbure-dimension. C'est cette approche que j'ai étendu au cadre de variétés riemannienne en améliorant la classe des torsions licites.

2 Le problème de Brödinger

Ce problème est une autre illustration du lien entre calcul stochastique et analyse. Brenier a fait un lien entre les probabilités minimisant une certaine énergie (problème de Brenier) et les solutions de l'équation d'Euler. Le problème de Brödinger - ou problème de Brenier-Schrödinger - est une adaptation du problème de Brenier à l'équation de Navier-Stokes. Il s'agit d'un problème de minimisation d'entropie avec des contraintes sur les marginales :

$$H(P|R) \to \min; [P_t = \mu_t, \forall t \in \mathcal{T}], P_{01} = \pi, \quad (12)$$

où R est une mesure de référence sur l'espace des chemins sur M (la mesure réversible du brownien par exemple), les μ_t sont des mesures sur l'espace M et π une mesure sur $M \times M$. Dans [3], il est montré que dans les cas où M et $\mathbb{R}^n$ ou le tore, les solutions de ce problème sont de mesure de semi-martingale dont la dérive vérifie l'équation de Burgers. Il est aussi démonter une CNS d'existence de solution à ce problème . Je travaille à étendre ces résultat dans le cadre des variétés. D'une part, je montre que pour les variétés à bord, les solutions sont aussi des mesure de semi-martingale dont la dérive vérifie l'équation de Burgers et est tangentielle au bord. D'autre part, je donne une CNS d'existence de solution pour les espaces symétriques.

References

[1] Marc Arnaudon, Michel Bonnefont, Aldéric Joulin. *Intertwinings and generalized Brascamp-Lieb inequalities*. Rev. Mat. Iberoam., 2018.

[2] Marc Arnaudon, Koléhè A. Coulibaly, Anton Thalmaier. *Horizontal Diffusion in C^1 Path Space*. Séminaire de Probabilités XLIII, 2011.

[3] Marc Arnaudon, Ana Cruzeiro, Christian Léonard, Jean-Claude Zambrini. *An entropic interpolation problem for imcompressible viscous fluids*. Annales de l'Institut Henri Poincaré, à paraître.

[4] Baptiste Huguet. *Intertwining relations for diffusions in manifolds and applications to functional inequalities*. Preprint

Numerical Methods to Solve The Electrocardiographic Inverse Problem

Amel KAROUI
Directeurs : Nejib ZEMZEMI, Mostafa BENDAHAMANE

Institute of Mathematics of Bordeaux,UMR 5251, University of Bordeaux

In the last decades, cardiac diseases remained the worldwide leading cause of death. Since then, the electrocardiography has been an active research area where researchers are interested in modeling the electrical activity of the heart in order to better understand the physical phenomenon with the aim of helping clinicians in diagnosing and why not curing cardiac pathologies. In this context, the objective of my thesis is to develop numerical tools to non-invasively collect heart electrical data using electrocardiograms(ECGs) and heart-torso geometry acquired from medical imaging. This is what we call the electrocardiography imaging inverse problem.

1 State-of-the-art

In the litterature, almost all of the works use a mathematical approach to model the propagation of the electrical activity between the heart and torso surfaces (respectively Γ_H and Γ_{ext}) based on the phenomenon physics. In this context, a wide range of mathematical models were written going from the most complex ones that take into account all the heterogeneities of the torso domain and the ionic transportations to simplified ones. One of the latters consists of modeling the forward relationship between extracellular electrical potential u_H on the heart surface and the electrical potential u_T in the torso domain Ω_T and specifically on the surface Γ_{ext} using the following Laplace equation:

$$\begin{cases} \nabla \cdot (\sigma_T \nabla u_T) = 0, & \text{in} \quad \Omega_T, \\ \sigma_T \nabla u_T \cdot n_T = 0, & \text{on} \quad \Gamma_{ext}, \\ u_T = u_H, & \text{on} \quad \Gamma_H. \end{cases} \tag{13}$$

where σ_T stands for the torso conductivity tensor and n_T is the outward unit normal to the torso external boundary Γ_{ext}. This formulation provides, with slight modifications, the mathematical model of the inverse problem. Many research studies came up with different approaches to solve such problems among which we cite: the finite element method (FEM), the boundary element method (BEM) and the method of fundamental solutions (MFS). Using any of these numerical approaches, the governing equation (13) can be reduced to a matrix-vector system :

$$Ax = b \tag{14}$$

where A is the transfer matrix, b is the known boundary condition vector and x is the unknown vector. Their forms depend on the chosen numerical method. Generally, the inverse problem of electrocardiography represented by (14) is ill-posed which means that a small perturbation of the boundary condition data b may lead to a high variation in the inverse solution x. To overcome this issue, a regularization approach is often used to solve (14). The inverse problem is henceforth written as follows :

$$\min_x \left\{ \|Ax - b\|^2 + \lambda^2 \|Lx\|^2 \right\} \tag{15}$$

where L is the regularization operator and λ is the regularization parameter. In this thesis, we work on improving state-of-the-art methods performance and suggest new approaches to solve the inverse problem of electrocardiography.

2 Contributions

The first part of the thesis focuses on evaluating many numerical approaches combinations in order to determine the most reliable method to solve the inverse problem in terms of epicardial electrical potential. It consists of assessing the performance of two numerical

methods (FEM and MFS), each combined with two different regularization operators (Zero order Tikhonov and L1-norm regularization of the current density over the heart surface) and five methods for choosing the regularization parameter (UCurve, Creso, ADPC, GCV and RGCV). More details can be found in [1]. This study allowed to build a benchmark of state-of-the-art methods used to solve the inverse problem of electrocardiography.
In the second part, the main goal is to explore data-based methods, especially artificial neural networks, to give a new solving approach of the inverse problem. To do so, we suggest using a time-delay neural network (TDNN) to reconstruct from body surface potentials (BSPs) the corresponding heart surface potentials (HSPs) at a given timestep of the heartbeat. The main idea is that the BSP at a timestep t is highly dependent with its values at previous timesteps $t-1$, $t-2$,... due to the propagation phenomenon. Thus, TDNN uses as input the BSP at timestep t and its n previous values where n is a parameter to fix to estimate the HSP at the same timestep t. Similarly to the temporal correlation, we suppose that HSP in a given point P is strongly dependent on its recorded values at the adjacent points. Hence, we suggest a spatial adaptation of TDNN where both temporal and spatial correlation are considered. This study was the subject of a publication [2].
The last part addresses the cardiac activation mapping problem. This is an important diagnostic tool to understand cardiac arrhythmias caused by an electrical disorder. It consists of generating a map of the depolarization sequence of a heartbeat. This map is constructed from HSPs. These latter signals are either recorded invasively through catheterization or non-invasively reconstructed using inverse methods. Our purpose is to evaluate the performance of artificial neural networks on reconstructing activation maps non-invasively directly from BSPs without providing the HSPs [3].
To make the links between the different parts of the thesis, a comparison study is performed between physics-based and data-based inverse methods in terms of activation mapping.

3 Results

Results of the first part show that solving the inverse problem using the L1-norm regularization combined with FEM and RGCV gives the best results in terms of accuracy of epicardial potential reconstruction and pacing sites localization.
The second part presents a proof-of-concept showing the ability of machine learning techniques to solve the electrocardiographic inverse problem. Numerical results show that the spatial adaptation of TDNN is more accurate than TDNN.
The third part presents also a proof-of-concept showing that neural networks are a good alternative to cardiac activation mapping. This is emphasized by the comparison study performed between three different approaches of cardiac activation mapping :

- using reconstructed HSPs by FEM combined with L1-norm regularization.
- using reconstructed HSPs by the spatial adaptation of TDNN
- directly from BSPs using artificial neural networks.

Results show that direct mapping outperforms the other methods.

4 Conclusion and perspectives

Each one of the studies presented above comes up with a novelty in comparison with the state-of-the-art. Nevertheless, they all have limitations that need to be addressed in future works. First, we need to validate all the methods on a clinical dataset collected on patients. Second, the challenge of standardizing the heart-torso geometry in order to have a generalized data-based model applicable for all patients remains current. Finally, data-based models would be more credible if built using more complex data that illustrate real cardiac diseases.

References

[1] A. Karoui, P. Migerditichan, L. Baer, M. Bendahamane, N. Zemzemi *Evaluation of fifteen algorithms for the resolution of the electrocardiography imaging inverse problem using ex-vivo and in-silico data*, Frontiers in Physiology, **9**,(2018) 1708.

[2] A. Karoui, M. Bendahamane, N. Zemzemi *A Spatial Adaptation of the Time Delay Neural Network for Solving ECGI Inverse Problem*, In International Conference on Functional Imaging and Modeling of the Heart, pp. 94-102. Springer, Cham, 2019.

[3] A. Karoui, M. Bendahmane, N. Zemzemi *Direct Mapping from Body Surface Potentials to Cardiac Activation Maps Using Neural Networks*, In 2019 Computing in Cardiology (CinC). IEEE, 2019.

Elaboration of large-scale brain network atlases underpinning cognitive functions from a functional neuroimaging database of 297 healthy subjects

Application to the study of inter-individual variability of language

Loïc LABACHE
PhD supervisors: Marc Joliot & Jérôme Saracco
CEA PhD supervisor: Nathalie Tzourio-Mazoyer

Groupe d'Imagerie Neurofonctionnelle – CEA & IMN, UMR 5293, Université de Bordeaux
Contrôle de Qualité et Fiabilité Dynamique – Inria Bordeaux Sud Ouest & IMB, UMR 5251, ENSC - Bordeaux INP

> *"A science of the mind must reduce... complexities (of behavior) to their elements. A science of the brain must point out the functions of its elements. A science of the relations of mind and brain must show how the elementary ingredients of the former correspond to the elementary functions of the latter."*[JAM90]

This is how, in 1890, William James, one of the founders of American psychology, foresaw the evolution of his emerging field of research. A century and thirty years later, within the neurosciences, constructing such correspondences between cerebral functional architecture (with neuroimaging) and cognitive psychology remains one of the great contemporary scientific challenges.

1 Introduction

My thesis work is part of this multi-modal and multi-scale integration approach which has led, over the last decade, to the emergence of cognitive neuroimaging and then, via the development of large databases, to population neuroimaging. More specifically, fMRI (functional Magnetic Resonance Imaging) provides two types of three-dimensional brain maps from the BOLD (Blood Oxygen Level Dependent) signal. The BOLD signal is an indirect measure of neural activity reflecting the input and intracortical processes taking place in a given brain area [LOG01].

The first type of maps, activation maps, allows for the visualization of brain regions directly involved in a cognitive task, while the second type, intrinsic connectivity maps, allows for the study of synchronization between spatially distant but functionally connected and synchronous regions that share information with each other [MAZ01].

My PhD thesis work relies on these two brain map modalities to which I applied new statistical methodologies allowing for treating both the individual and the spatial dimensions.

In a first part, I designed atlases of brain regions dedicated to specific cognitive functions based on their hemispheric lateralization and targeting the average results of populations selected for their low variability. In a second part, I studied the inter-individual variability of the functional brain organization of the individuals and networks involved.

To do this, I was able to use data from the BIL&GIN database (Brain Imaging of Lateralization by the Groupe d'Imagerie Neurofonctionnelle: [MAZ16]). The BIL&GIN is a multimodal database specifically conceived for the study of hemispheric specialization, which is a hemispheric interdependent relationship between a motor, sensory or cognitive function and a set of brain structures. This interdependence includes both the localization within a given hemisphere of the networks that have unique functional properties and the mechanisms that enable the inter-hemispheric coordination necessary for an efficient processing [TZO16]. The BIL&GIN was acquired to gain understanding on the role of hemispheric dominance in the establishment of large-scale brain organization.

I started by developing supra-modal atlases of the processing areas at the word level [HES19] and at the sentence level [LAB19] in relation to the hierarchical organization of language networks.

2 Anatomofunctional organization of language

Indeed, although there are many approaches to map language areas in patients, most often in the context of preoperative (Wada test or electrocorticographic recording of cognitive evoked potentials)) or intraoperative evaluation (cortical stimulation), there is no atlas of the networks supporting language functions in healthy subjects yet.

To define the areas involved in language processing in healthy subjects, I used two major characteristics of language organization in the brain. The first one is that language is hosted by the left hemisphere in 98% of right-handed individuals [MAZ14]. The second one is that the so-called essential language areas (the regions which lesions lead to an impairment of language) are recruited independently of the modality in which the language is processed.

2.1 SENtence Supramodal Areas AtlaS

In a first step, I identified the regions that were activated in the left hemisphere and leftward asymmetrical during sentence production, listening, and reading (conjunction approach) in 137 right-handed subjects.

Actually, 32 multimodal left asymmetrical and activated brain areas during sentence production, reading and listening were identified and grouped into a functional brain atlas called SENSAAS (SENtence Supramodal Areas AtlaS) [LAB19]. The temporal correlations at rest between these 32 regions made it possible to detect that they belonged to 3 networks. Among these networks, one including 18 regions, contains the essential areas of language, i.e. the areas whose lesion would cause an impairment in the integration of the meaning of speech.

In a second step, in order to further study this network, I used graph theory tools to characterize each of the 18 regions in terms of degree centrality and sunnexity. It turned out that the anterior and posterior poles of this network were hubs; both in terms of information integration but also in terms of information diffusion through the network. This antero-posterior pole really corresponds to the epicentres of language whose lesion leads to aphasia [MES00].

This atlas has already been used in cohort studies looking for the genetic basis of language regions [CAR19] or the neuroanatomical basis of their pathology, such as in the study of the consequences of temporal epilepsy on language and memory networks [ROG20].

2.2 Variability of language dominance

In a third step, I used SENSAAS to study the inter-individual variability of hemispheric lateralization of language in 287 BIL&GIN individuals including 144 left-handed people [LAB20a].

To do so, I used an agglomerative hierarchical classification method including several modalities and several scales. Indeed, each of our partitions are not only characterized by asymmetries during the three language tasks (producing, listening and reading) but also by degree centrality and inter-hemispheric correlation, both at the level of hubs and networks. This work identifies three groups of individuals characterized by different inter- and intra-hemispheric organizations of language network: two groups of typical subjects and one group of atypical subjects.

It also allowed us to characterize in these 3 groups the variability of the integration of information at rest in the SENSAAS regions, to analyze the occurrence of lateralization dissociations during the different language modalities and to show differences in anatomical organization and cognitive skills.

2.3 Word-list Multimodal Cortical Atlas

I applied the same two-steps method to define an atlas of the word processing networks [HES19], 21 brain regions were identified and appeared to be organized in 2 distinct network.

A network in particular, constituted of strong connections between perceptive and action areas, was identified as supporting phonological processing. Interestingly, within this network a strong connectivity was observed between the audio-motor loop of the left hemisphere and the right superior temporal sulcus supporting prosodic processing.

3 Similarity of Individual MatriceS

Finally, I developed a new method to directly study the variability of the organization of intrinsic connectivity using correlation matrices and an agglomerative hierarchical classification algorithm [LAB20b], based on the work of Suzuky [SUZ06], which allows us to study the stability of partitions based on the empirical frequency of partitions occurrence across the population and not on the mean intrinsic connectivity matrix.

I proposed two different mathematical tools allowing, for the first one, for identifying the network or brain region leading to partition instability when using a hierarchical ascending classification method, and for the other one, for extracting stable sub-populations from a starting population, based on the inter-individual variability of their dendrogram.

References

[CAR19] Carrion-Castillo A., Van der Haegen L., Tzourio-Mazoyer N., Kavaklioglu T., Badillo

T., Chavent M., Saracco J., Brysbaert M., E. Fisher S., Mazoyer B., Francks C. *Genome sequencing for rightward hemispheric language dominance*. Genes, Brain and Behavior 18.5 (2019): e12572.

[HES19] Hesling I., **Labache L.**, Joliot M., Tzourio-Mazoyer N. *Large-scale plurimodal networks common to listening to, producing and reading word lists: an fMRI study combining task-induced activation and intrinsic connectivity in 144 right-handers*. Brain Structure and Function 224.9 (2019): 3075-3094.

[JAM90] William J. *The principles of psychology*. Vol. 1. No. 2. London: Macmillan, 1890.

[LAB19] **Labache L.**, Joliot M., Saracco J., Jobard G., Hesling I., Zago L., Mellet E., Petit L., Crivello F., Mazoyer B., Tzourio-Mazoyer N. *A SENtence Supramodal Areas AtlaS (SENSAAS) based on multiple task-induced activation mapping and graph analysis of intrinsic connectivity in 144 healthy right-handers*. Brain Structure and Function 224.2 (2019): 859- 882.

[LAB20a] **Labache L.**, Joliot M., Saracco J., Tzourio-Mazoyer N. *Characterization of typical and atypical brain organization for language through a joint analysis of intrinsic connectivity strength and task-induced functional asymmetries during sentence production, listening and reading*. Submitted to Cerebral Cortex (2020).

[LAB20b] **Labache L.**, Joliot M., Saracco J. *Similarity of Individual MatriceS; detection of unstable networks and sub-populations in a three-dimensional table*. In preparation (2020).

[LOG01] K. Logothetis N., Pauls J., Augath M., Trinath T., Oeltermann A. *Neurophysiological investigation of the basis of the fMRI signal*. Nature 412.6843 (2001): 150-157.

[MAZ01] Mazoyer B., Zago L., Mellet E., Bricogne S., Etard O., Houdé O., Crivello F., Joliot M., Petit L., Tzourio-Mazoyer N. *Cortical networks for working memory and executive functions sustain the conscious resting state in man*. Brain research bulletin 54.3 (2001): 287- 298.

[MAZ14] Mazoyer B., Zago L., Jobard G., Crivello F., Joliot M., Perchey G., Mellet E., Petit L., Tzourio-Mazoyer N. *Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness*. PloS one 9.6 (2014).

[MAZ16] Mazoyer B., Mellet E., Perchey G., Zago L., Crivello F., Jobard G., Delcroix N., Vigneau M., Leroux G., Petit L., Joliot M., Tzourio-Mazoyer N. *BIL&GIN: a neuroimaging, cognitive, behavioral, and genetic database for the study of human brain lateralization*. Neuroimage 124 (2016): 1225-1231.

[MES00] Mesulam M. *Principles of behavioral and cognitive neurology*. Oxford University Press, 2000.

[ROG20] Roger E., Pichat C., Torlay L., David O., Renard F., Banjac S., Attyé A., Minotti L., Lamalle L., Kahane P., Baciu M. *Hubs disruption in mesial temporal lobe epilepsy. A resting-state fMRI study on a language-and-memory network*. Human brain mapping 41.3 (2020): 779- 796.

[SUZ06] Suzuki R., Hidetoshi S. *Pvclust: an R package for assessing the uncertainty in hierarchical clustering*. Bioinformatics 22.12 (2006): 1540-1542.

[TZO16] Tzourio-Mazoyer N. *Intra-and inter-hemispheric connectivity supporting hemispheric specialization*. Micro-, meso-and macro-connectomics of the brain. Springer, Cham, 2016. 129- 146.

Spaces of analytic functions and Reachable space of the heat equation

Marcu-Antone ORSONI
Directeur : Andreas HARTMANN

Institut de Mathématiques de Bordeaux, CNRS UMR 5251, Université de Bordeaux

1 Heat equation and initial control problem

Control theory studies the possibility of acting on a dynamical system by means of a command, also called control. A good example of such a system is given by the one-dimensional heat equation on a segment with Dirichlet boundary control

$$\begin{cases} \dfrac{\partial w}{\partial t}(t,x) - \dfrac{\partial^2 w}{\partial x^2}(t,x) = 0 & t>0,\ x\in(0,\pi), \\ w(t,0) = u_0(t), \quad w(t,\pi) = u_\pi(t) & t>0, \\ w(0,x) = f(x) & x\in(0,\pi). \end{cases} \tag{HE}$$

It models the evolution of the temperature w on a rod depending on the time t when we heat the rod at both ends. Here we act on the system by means of the boundary heating function $u := (u_0, u_\pi)$, so called control, at each instant t. This system is well-posed: for any control $u := (u_0, u_\pi) \in L^2_{\text{loc}}((0,+\infty),\mathbb{C}^2)$ and any initial condition $f \in X := W^{-1,2}(0,\pi)$ (the dual of the Sobolev space $W^{1,2}_0(0,\pi)$), this equation admits a unique solution $w \in C([0,+\infty), X)$.

If for any initial state $f \in X$ and any final target $g \in X$ we can find a control $u \in L^2((0,T),\mathbb{C}^2)$ in order to steer the temperature from the initial state f (at time $t=0$) to the final target g (at time $t=T$), then we say that the system is (exactly) controllable (at time T). Unfortunately, this controllability property is not fulfilled by our heat system. This is a consequence of the extreme regularity in the space variable x of the temperature.

We need some definitions. For $T>0$, we say that $g \in X$ is reachable in time $T>0$ if there exists a boundary control $u \in L^2((0,T),\mathbb{C}^2)$ such that the solution of (HE) satisfies $w(T,\cdot) = g$. We denote by $\mathcal{R}^f_T$ the set of all reachable functions from f in time T. This set $\mathcal{R} := \mathcal{R}^f_T$ is known to depend neither on the initial condition $f \in X$, nor on time T and we have seen in the previous paragraph that it is strictly contained in X. This leads to the natural question: which are exactly the targets we can reach? The aim of my thesis is to give an exact and definitive characterization of the reachable space $\mathcal{R}$.

2 The reachable space as a space of analytic functions

The description of the space $\mathcal{R}$ is a central question in control theory which goes back to the work [2] of Fattorini and Russell, 50 years ago. It has a long history of results with different methods and has gained quite some renewed interest in recent years. Actually, it is a folklore result of PDEs that the reachable states g extend holomorphically on the square D one diagonal of which is the interval $(0,\pi)$ (see Figure 1):

$$D = \left\{ z = x+iy \in \mathbb{C} \mid \left| x - \frac{\pi}{2} \right| + |y| < \frac{\pi}{2} \right\}.$$

Conversely, Dardé and Ervedoza showed in [1] that for every neighborhood D_ε of D, the analytic functions on D_ε are reachable. Hence, $\mathcal{R}$ is a space of analytic functions on D which we have to determine.

Before my PhD thesis, the best known result on this problem was given in [3], where Hartmann, Kellay and

Tucsnak proved that the reachable space is sandwiched between two well-known Hilbert spaces of holomorphic functions on the square D. They also conjectured that it is equal to the Bergman space of the square D:

$$\mathcal{R} = A^2(D). \qquad \text{(HKT-conjecture)}$$

For an open connected set Ω in $\mathbb{C}$, the Bergman space $A^p(\Omega)$ $(1 \leq p < \infty)$ consists of all the functions which are holomorphic and L^p on Ω.

In my first work [5], I proved that $\mathcal{R}$ is the sum of two Bergman spaces on sectors the intersection of which is the square D (see Figure 1):

Theorem 2. *We have* $\mathcal{R} = A^2(\Delta) + A^2(\pi - \Delta)$ *where* $\Delta = \left\{z \in \mathbb{C}, |\arg(z)| < \frac{\pi}{4}\right\}$.

This has been achieved using a Paley-Wiener type theorem for the Bergman space.

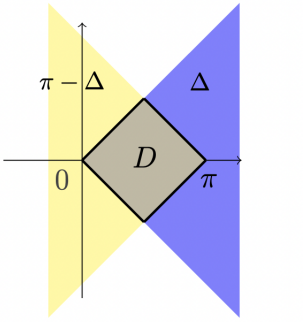

Figure 1: *The square D as intersection of the sectors Δ and $\pi - \Delta$.*

3 Separation of singularities and final solution

If we want to prove the (HKT-conjecture), by Theorem 2 it remains to show that the Bergman space $A^2(D)$ identifies with the sum $A^2(\Delta) + A^2(\pi - \Delta)$. It is important to notice that the square D is precisely the intersection of the sectors Δ and $\pi - \Delta$ (see Figure 1). We will now solve a more general problem and obtain as a consequence the required equality.

For $\Omega \subset \mathbb{C}$, an open set in the complex plane, we denote by $\mathrm{Hol}(\Omega)$ the space of holomorphic functions on Ω. Given Ω_1 and Ω_2 two open subsets of $\mathbb{C}$ with non empty intersection, a natural question is to know whether every function $f \in \mathrm{Hol}(\Omega_1 \cap \Omega_2)$ can be written as a sum of two functions $f_1 \in \mathrm{Hol}(\Omega_1)$ and $f_2 \in \mathrm{Hol}(\Omega_2)$, i.e. is the equality $\mathrm{Hol}(\Omega_1 \cap \Omega_2) = \mathrm{Hol}(\Omega_1) + \mathrm{Hol}(\Omega_2)$ true? This problem is known as the separation of singularities problem for holomorphic functions and has a quite long history. Aronszajn gave a positive answer for any pair (Ω_1, Ω_2) of open sets in $\mathbb{C}$.

Since then, the problem has attracted a lot of interest in particular in Banach spaces of analytic functions (e.g. for bounded analytic functions). Here we are interested in separation of singularities for the Bergman spaces, i.e. if $1 < p < \infty$, does the equality $A^p(\Omega_1 \cap \Omega_2) = A^p(\Omega_1) + A^p(\Omega_2)$ hold? We showed in [4] that the answer is positive in several settings, including sectors or polygons and quite general bounded convex sets.

Theorem 3. *Let $1 < p < \infty$. Let Ω_1 and Ω_2 be two open convex sets in $\mathbb{C}$ such that*

(i) $\Omega_1 \cap \Omega_2$ is non-empty and bounded,
(ii) The set $\partial\Omega_1 \cap \partial\Omega_2$ has a finite (possibly zero) number of connected components.

Then $A^p(\Omega_1 \cap \Omega_2) = A^p(\Omega_1) + A^p(\Omega_2)$.

As a consequence we obtained the following result.

Corollary 4. *The equality $A^2(D) = A^2(\Delta) + A^2(\pi - \Delta)$ holds.*

This proves the (HKT-conjecture) and gives a definitive answer to the 50 years old problem of determining the reachable space of the heat equation.

References

[1] J. Dardé and S. Ervedoza. *On the reachable set for the one-dimensional heat equation*. SIAM J. Control Optim., 3:1692–1715, 2018.

[2] H.O. Fattorini and D.L. Russell. *Exact controllability theorems for linear parabolic equations in one space dimension*. Arch. Rational Mech. Anal., pages 272–292, 1971.

[3] Andreas Hartmann, Karim Kellay, and Marius Tucsnak. *From the reachable space of the heat equation to Hilbert spaces of holomorphic functions*. J. Eur. Math. Soc. (JEMS), 22(10):3417–3440, 2020.

[4] Andreas Hartmann and Marcu-Antone Orsoni. *Separation of singularities for the Bergman space and application to control theory*. J. Math. Pures Appl. (9), 150:181– 201, 2021.

[5] Marcu-Antone Orsoni. *Reachable states and holomorphic function spaces for the 1-D heat equation*. J. Funct. Anal., 280(7):108852, 17, 2021.

Random Tensor and the Sachdev-Ye-Kitaev model

Romain PASCALIE
Directeurs : Adrian TANASA, Raimar WULKENHAAR

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux
Mathematisches Institut der Westfalischen Wilhelms-Universitaat, Université de Münster

Tensor models are a generalisation of matrix models introduced to study triangulisations of spaces of dimension bigger than 2. A particular class of graphs, called melonic, dominates these models in the so-called large N limite, N being the size of the tensor. On one hand, we studied a particular class of tensor models using quantum field theoretical techniques. On the other hand we studied some mathematical aspects of the Sachdev-Ye-Kitaev (SYK) model, which is a toy model of high energy physics.

1 Quantum field theory and Matrix models

Quantum field theory is a general framework used in theoretical and mathematical physics to study vairous physical models. Starting from a generating function which fixes the specificity of the model under study, one is able to compute so called correlation function which encode the physical behaviour of the model. An example of a generating function in the case of matrix models is

$$Z = \int_{\mathbb{H}_N} \mathrm{d}M \exp\Big(-\frac{1}{2}\mathrm{Tr}(M^2) + \frac{\lambda}{3}\mathrm{Tr}(M^3)\Big), \quad (16)$$

where λ is called the coupling constant and the integration is on $N \times N$ Hermitian matrices with the measure $\mathrm{d}M = \prod_{a,b=1}^{N} \mathrm{d}M_{ab}$. The correlation function are then defined as an expectation value of some quantities built from the matrix M (which is called the field), for instance:

$$< \mathrm{Tr}(M^k) >= \int_{\mathbb{H}_N} \mathrm{d}M\, \mathrm{Tr}(M^k) exp\Big(-\frac{1}{2}\mathrm{Tr}(M^2) + \frac{\lambda}{3}\mathrm{Tr}(M^3)\Big)$$
$$= \sum_{n\geq 0} \frac{\lambda^n}{n!} \int_{\mathbb{H}_N} \mathrm{d}M\, \mathrm{Tr}(M^k)\mathrm{Tr}(M^3)^n exp\Big(-\frac{1}{2}\mathrm{Tr}(M^2)\Big).$$

Using a series expansion of the quartic term of the exponential (called interaction term), one can compute order by order in λ the correlation function. At each order in λ, the computation of the integral is done using Wick theorem which performs pairings of the different terms in the integrand. Each combination of pairing can be represented as a graph using a set of rules which depends on the model. These graphs are the so-called Feynman graphs. In the language of probabilities, the correlation functions correspond to the moments of the probability distribution, or cumulants if we consider only connected graphs.

In the case of matrix model, the Feynman graphs are ribbon graphs (see Fig. 1). An important result in matrix models is that when the size of the matrix N tend toward infinity, only the planar graph dominates and the other graphs are suppressed.

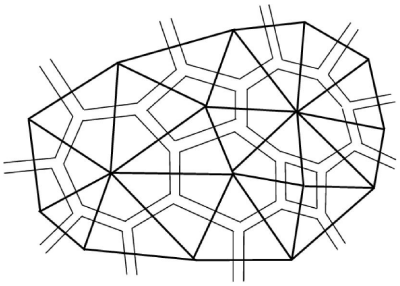

Figure 1: *Triangulation of a 2D surface and its dual, a ribbon graph.*

Matrix models are related, amongst other fields, to 2D quantum gravity. This relation is based on a representation of discrete gravity in terms of random matrices (see, for example, the review article of Di Francesco et. al. [1]), representation which requires, from a mathematical point of view, to discretise surfaces. The discretisation of 2D surfaces is done in terms

of triangulations. One can now take a dual point of view and encode the information coming from this triangulation in a ribbon graph picture (see Fig. 1).

Matrix models are an essentially 2D tool, their natural generalisation to dimensions higher than two are the random tensor models.

2 Tensor models

Ribbon graphs can be generalised to tensor graphs, which are graphs associated with random tensor models [2].

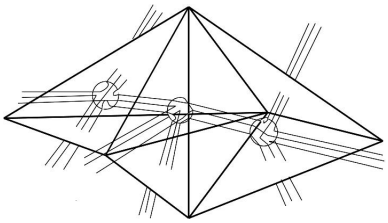

Figure 2: *"Triangulation" of a 3D space. The dual is a rank three tensor graph.*

In the three-dimensional case, the tri-valent vertices from above become valence four vertices. Such a vertex corresponds to a tetrahedron, the simplest building-block of a three-dimensional space, thus generalising the image of the two-dimensional case. This is represented in Fig. 2.

When the size N of the tensor tend to infinity, the dominating graphs are now a particular class of graphs called melonic graphs (see Fig. 3).

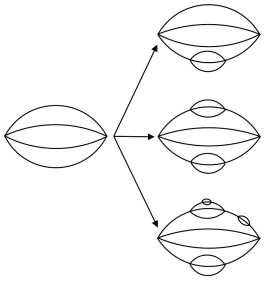

Figure 3: *Melonic graphs - they are obtained from insertions of melonic subgraphs.*

3 The Sachdev-Ye-Kitaev model

Another aspect of my work concerns the SYK model. It was proposed by Kitaev as a toy model of holography [3] and has attracted a huge amount of interest within the AdS/CFT community of high energy physics due to its very particular properties. The model was studied in [4] and many other papers. A crucial fact is that the leading graph of this model are the melonic graph, as in tensor models. This was pointed out by Witten [5], who proposed a reformulation of the SYK model, in form of a tensor model.

4 Results

In my PhD, we have studied a particular type of tensor models by generalising techniques from matrix models. We obtained a system of coupled equations for the correlation functions, called Schwinger-Dyson equations [6] where each correlation function is indexed by a graph and the equations are obtained in part by graph manipulations.

We then simplified this system of equation by defining the appropriate large N limit [7], which partially decouples the system.

Furthermore, we solved the system for a simple model [8] using perturbation expansion and Lagrange-Bürmann formula.

In the original SYK model the coupling constant is a random tensor which is averaged under a Gaussian distribution. We studied the effect of a non-Gaussian distribution on the physics of the model, showing that at leading order in the large N limit, only the Gaussian behaviour dominates [9].

References

[1] P. Di Francesco, P. H. Ginsparg and J. Zinn-Justin, Phys. Rept. **254** (1995) 1 doi:10.1016/0370-1573(94)00084-G

[2] R. Gurau, "Random Tensor", Oxford University Press, 2017

[3] A. Kitaev, Talks at KITP, April 7, and May 27, 2015 http://online.kitp.ucsb.edu/online/entangled15/kitaev http://online.kitp.ucsb.edu/online/entangled15/kitaev2

[4] J. Maldacena and D. Stanford, Phys. Rev. D **94** (2016) no.10, 106002 doi:10.1103/PhysRevD.94.106002

[5] E. Witten, J. Phys. A **52** (2019) no.47, 474002 doi:10.1088/1751-8121/ab3752

[6] R. Pascalie, C. I. P. Sánchez and R. Wulkenhaar, accepted in Ann. Henri Poincaré D arXiv:1706.07358 [math-ph].

[7] R. Pascalie, C. I. Pérez-Sánchez, A. Tanasa and R. Wulkenhaar, J. Math. Phys. **60** (2019) no.7, 073502 doi:10.1063/1.5080306

[8] R. Pascalie, Lett. Math. Phys. (2019) doi:10.1007/s11005-019-01245-0

[9] T. Krajewski, M. Laudonio, R. Pascalie and A. Tanasa, Phys. Rev. D **99** (2019) no.12, 126014 doi:10.1103/PhysRevD.99.126014

Modelling and simulating the zebrafish escape swimming: an experiment-driven approach

Guillaume RAVEL
Co-Directrice : Afaf BOUHARGUANE
Co-Directeur : Patrick J. BABIN
Collaborateurs : Anja KNOLL-GELLIDA, Michel BERGMANN, Angelo IOLLO

Institut de Mathématiques de Bordeaux, UMR 5251, Université de Bordeaux
Equipe-projet Memphis, Inria Bordeaux Sud-Ouest
Laboratoire Inserm Maladies Rares : Génétique et Métabolisme, Inserm U1211, Université de Bordeaux

When stimulated, zebrafish eleutheroembryos/larvae perform the so-called C-start escape response, a very stereotyped behaviour. In the biology field, only swimming kinematics were approximated to analyse the escape behaviour and compare different zebrafish swimming altered by drugs. Biologists use the zebrafish as an animal model to study the effects of neurotoxic drugs on the locomotion and develop pharmacological treatments. Here, we aimed at performing numerical simulations to reproduce the experimental behaviour and enhance the understanding of the zebrafish eleutheroembryo swimming, while visualizing the fluid flows surrounding the zebrafish. The computational model which has been developed, includes the recovery of the experimental deformation velocity field paired with a computational fluid dynamic environment. The innovative aspect resides in computing the deformation velocity field from experimental data. Thus, the computational fluid dynamics model provides the full 3D computation of the flow and gives an insight of the energetic performance during the escape swimming.

1 Context

Swimming simulations are performed to study the performance of an active immersed body according to its shape and deforming movements. The deformation velocity of the body is usually imposed using a proper analytic formulation to model simple deforming movements. In the literature, more complex motions can be derived to model food and prey captures [1]. To our knowledge, only few studies tried to extract the body deformations directly from actual data, using the optimal transport theory [2] or experimental observations [3]. Here, we aim at modelling a real-like 3D eleutheroembryo zebrafish shape and applying the actual deformation extracted from experimental images to simulate the zebrafish escape response. The deformation is recovered using either optimal transport theory or the reconstruction of 3D deformed surfaces, depending on the number of dimensions considered.

Our numerical model is based on a Cartesian mesh and immersed boundary methods to solve penalised Navier-Stokes equations. In the end, the simulation enlightens the amount of energy expanded by the zebrafish during the locomotion for any experimental conditions. To validate the model, we recorded multiple experimental videos of the locomotion after electric field pulse stimulation and challenged the escape response by modifying the fluid dynamic viscosity or after drug exposures. Experimental and numerical trajectories have been successfully compared.

2 Numerical Simulations

2.1 The CFD method

We approximated the incompressible Navier-Stokes equations on a Cartesian mesh using a penalisation

method combined with a level-set method (see [1] for more details). Velocities are computed solving penalised Incompressible *Navier-Stokes* equations where χ_s is the body characteristic function and $\mathbf{u_s}$ is the body velocity.

1. Penalised *Navier-Stokes* Equations ($\lambda \gg 1$):

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu\Delta\mathbf{u} + \chi_s\lambda(\mathbf{u_s} - \mathbf{u}) \quad (17)$$

$$\nabla\cdot\mathbf{u} = 0 \quad (18)$$

2. $\mathbf{u_s}(\mathbf{x},t) = \bar{\mathbf{u}}(\mathbf{x},t) + \mathbf{u}^{\theta}(\mathbf{x},t) + \mathbf{u_{def}}(\mathbf{x},t)$

Rigid velocities (translation $\bar{\mathbf{u}}$ and rotation $\mathbf{u}^{\theta}$) are computed solving *Newton*'s laws while the deformation velocity $\mathbf{u_{def}}$ still needs to be approximated.

2.2 The 2D case

Deformation velocities have been extracted using Procrustes Analysis methods and optimal transport theory in the two-dimensional case.

- How to find the deformation velocity field between two consecutive frames ?

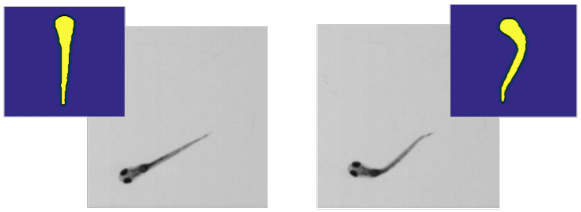

Figure 1: *Position at time t (left) and time $t+dt$ (right) Gray images - experimental frames; Binary images - recentered frames*

- *Monge-Kantorovich* formulation, from the optimal transport theory:

$$\begin{cases} \inf_X \int_{\Omega_0} \rho_0(\xi)|X(\xi)-\xi|^2 d\xi \\ \int_{\Omega_0}\rho_0(\xi)d\xi = \int_{\Omega_1}\rho_1(x)dx \end{cases} \quad (19)$$

Densities ρ_0 and ρ_1 are represented by the binary images (regularized) and the deformation velocity $X(\xi)-\xi$ is computed from *Picard* iterations (see [4] for more details).

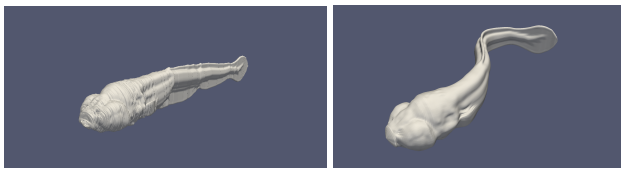

Figure 2: *Initial 3D zebrafish eleutheroembryo reconstruction (left) - Body-fitted surface of a deformed 3D reconstruction (right)*

2.3 The 3D model

For the three-dimensional case, we reconstructed a full 3D model from 18.207 transverse slices of a 5 days post-fertilization zebrafish [5], via a *Sinkhorn* algorithm before deforming it to provide a 3D surface mesh according to the midline deformations, as shown in Fig.2. Thus, deformation velocities are directly computed from surface meshes.

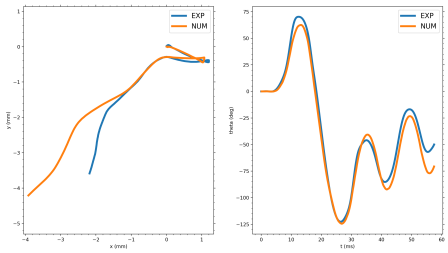

Figure 3: *Experimental and numerical kinematics* trajectory (left) and rotation (right)

As a result, we have been able to successfully reproduce multiple complex zebrafish escape swimming after computing experimental deformations. We can see in Fig. 3 an example of validation. The numerical simulation computes the whole fluid flow as illustrated in Fig.4. The computational model was used to quantify the power expanded by the zebrafish depending on its deformation.

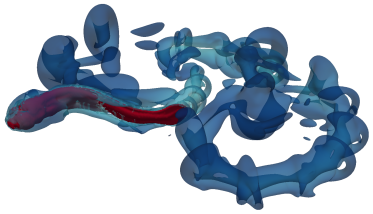

Figure 4: *Vorticity structure (iso-contour of q-criterion) coloured by velocity surrounding the fish body (red)*

References

[1] Bergmann, M., Iollo, A., *Modeling and simulation of fish-like swimming*, J.Comput.Phys., 2011.

[2] Bergmann, M., Iollo, A., *Bioinspired swimming simulations*, J.Comput.Phys., 2016.

[3] Li, G., Muller, U. K., van Leeuwen, J. L. and Liu, H., *Body dynamics and hydrodynamics of swimming fish larvae: a computational study*, J.Exp.Biol., 2012.

[4] Bouharguane, A., Iollo, A., Weynans, L., *Numerical solution of the Monge-Kantorovich problem by density lift-up continuation*, ESAIM: Math.Model.Num., 2015.

[5] Hildebrand et al, *Whole-brain serial-section electron microscopy in larval zebrafish*, Nature, 2017.

Reduced-order models: convergence between HPC and data for fluid mechanics

Sébastien RIFFAUD
Supervisor : Angelo IOLLO

Institut de Mathématiques de Bordeaux, CNRS UMR 5251, Université de Bordeaux

The democratization of HPC makes possible the numerical modelling of complex multi-physical phenomena that were up to now inaccessible, as well as the management and analysis of the mass of associated data. The reduced-order models aim to reconcile the modelling approach and the data approach. The basic idea is to determine from a learning database a mathematical model involving a small number of degrees of freedom, ie a 'reduced model'. The reduced modelling of a physical problem represents a compromise between speed of solution and precision of the result, ideally with rigorous quantitative guarantees on error boundaries.

1 Introduction

In numerical simulations, the dynamic of complex fluid flows are described by mathematical models involving the resolution of parametric, time-dependent, system of partial-differential equations. Since most of theses equations does not have analytic solution, numerical methods are used to transform the continuous solution into its discrete counterpart, leading to the resolution of a $N \times N$ system. However in many industrial applications, the computational complexity of the resulting system poses challenge du to its large number of degrees of freedom $N \approx O(10^6)$.

Reduced-order modelling aims to decrease the complexity of this system by reducing the number of unknowns $M \approx O(10^1)$. Instead of discretizing the solution without any knowledge of the system behaviour, the reduced-order model uses a posteriori information to significantly reduce the computational complexity. To this end, they consist of a computationally expensive training stage where solutions of the equation are acquired to learn the system behaviour and of an inexpensive simulation stage for new predictions. This construction of the reduced-order model contains machine learning approaches to achieve dimensionality reduction.

2 Governing equation

In this work, we consider parametric, time-dependent, system of partial-differential equations. After space semi-discretization, we obtain the large-scale $N \times N$ system

$$\frac{\partial u_h}{\partial t} = f(u_h, t; \mu) \tag{20}$$

where μ denotes the parameters (like the Reynolds number, the initial conditions or the boundary conditions as example), $u_h(\cdot, t; \mu) \in \mathbb{R}^N$ denotes the discrete solution and $f(u_h(\cdot, t; \mu), t; \mu) \in \mathbb{R}^N$ denotes the discretization of the spatial operator (for example by finite-difference, finite-element or finite-volume schemes). The system (20) of ordinary equations can then be solved in time by a multistep method or a Runge-Kutta scheme.

3 Reduced-order model

In the reduced-order model, the discrete solution is approximated by

$$\widetilde{u}_h(x, t; \mu) = \sum_{n=1}^{M} a_n(t; \mu)\Phi_n(x) \tag{21}$$

where the space functions $\Phi_n \in \mathbb{R}^N$ are built during the training stage and the time coefficient $a_n(t;\mu) \in \mathbb{R}$ are computed in the simulation stage. In this way, the approximate solution $\widetilde{u}_h(\cdot,t;\mu) \in \mathbb{R}^N$ is sought in the M-dimensional subspace spanned by the space functions, also known as the approximation space.

3.1 Training stage

In the training stage, the discrete solution is sampled in order to learn its behaviour over time and parameters. To this end, snapshots $s_j \in \mathbb{R}^N$ of the discrete solution are collected at different timesteps t_k and for different parameters μ_i (i.e. $s_j(x) = u_h(x, t_{k(j)}; \mu_{i(j)})$), leading to the snapshot database

$$S = \{s_1, s_2, \dots, s_K\} \tag{22}$$

The approximation space is then obtained by Proper Orthogonal Decomposition (POD) [1, 4]. The POD finds the optimal low-dimensional subspace which best represents the snapshot database to achieve dimensionality reduction. More precisely, the resulting approximation space is the subspace of rank M minimizing in the least-squares sense the difference between the snapshots and their projections onto this subspace

$$\begin{cases} \underset{\Phi_1,\dots,\Phi_M}{\text{minimize}} & \displaystyle\sum_{j=1}^{K} \int_\Omega \big(s_j(x) - \mathcal{P}_j(x)\big)^2 \,\mathrm{d}x \\ \text{subject to} & \displaystyle\int_\Omega \Phi_n(x)\Phi_m(x)\,\mathrm{d}x = \delta_{n,m} \end{cases} \tag{23}$$

where $\delta_{n,m}$ denotes the Kronecker delta and $\mathcal{P}_j \in \mathbb{R}^N$ denotes the projection of the snapshot s_j onto the subspace spanned by the space functions $\Big(\text{i.e. } \mathcal{P}_j(x) = \displaystyle\sum_{n=1}^{M} \Big(\int_\Omega s_j(y)\Phi_n(y)\,\mathrm{d}y\Big)\Phi_n(x)\Big)$. In practice, a small number M of space functions is sufficient to obtain accurate approximation of the snapshots, allowing to significantly reduce the number of unknowns (i.e. $M \ll N$).

3.2 Simulation stage

Once the approximation space is built, the approximate solution depends only on M time coefficients during the simulation stage. Inserting the approximate solution in the semi-discretized system (20) leads to the residual

$$\mathcal{R}_h(x,t;\mu) = \frac{\partial \widetilde{u}_h}{\partial t} - f(\widetilde{u}_h, t; \mu) \tag{24}$$

which does not necessarily vanish for the approximate solution. A popular method to compute the time coefficients is the Galerkin method where the residual is enforced to be orthogonal to the approximation space. By projecting the residual onto the space functions, we obtain the small-scale $M \times M$ system

$$\frac{da_n}{dt} = \int_\Omega f(\widetilde{u}_h, t; \mu)\Phi_n(x)\,\mathrm{d}x \tag{25}$$

where we use the orthonormality of the space functions. In this way, the computational complexity of the resulting system (25) is significant smaller than the one of the large-scale system (20), allowing fast or real-time simulations for new parameters.

4 Organization of the thesis

The accuracy of the reduced-order model can vary from very high in the limit of $M = N$ space functions, to very low if a too small number of space functions is used. The reduced-order model can therefore be view as a trade-off between accuracy and computational cost.

This thesis is divided in two parts. In the first one, a fast and accurate reduced-order model, applied for the first time to the BGK equation [3] (an approximation of the Boltzmann equation) was developed in [2]. In the second part, a new stabilization method of projection-based model order reduction was proposed for convection-dominated flow when dealing with shock waves and discontinuities.

References

[1] G. Berkooz, P. Holmes, J.L. Lumley, *The proper orthogonal decomposition in the analysis of turbulent flows.* Annual review of fluid mechanics 25, 539–575, 1993.

[2] F. Bernard, A. Iollo, and S. Riffaud. *Reduced-order model for the BGK equation based on POD and optimal transport.* Journal of Computational Physics, 373:545–570, 2018.

[3] P.L. Bhatnagar, E.P. Gross, M. Krook, *A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems*, Phys. Rev. 94 (3) (1954) 511.

[4] L. Sirovich, *Turbulence and the dynamics of coherent structures. i. coherent structures.*, Quarterly of applied mathematics 45, 561–571, 1987.

Improving the efficacy of nanoparticles in cancer therapy with data-driven mathematical modeling

Cristina VAGHI[1]
Directors: Sébastien BENZEKRY[1]**, Clair POIGNARD**[1]**, Raphaëlle FANCIULLINO**[2]

[1] *MONC, Inria Sud Ouest Bordeaux; Insitut de Mathématiques de Bordeaux CNRS UMR5251, France.*
[2] *SMARTc, CRCM Aix-Marseille University, France.*

Nanoparticles for drug delivery are a promising technology in cancer treatment, as they target tumor cells while preserving healthy tissues from the toxicity of the chemotherapy. However, the dynamic of penetration of these nanoparticles in tumor tissues is not well known. Mathematical modeling is helpful in order to quantify the absorption of the nanoparticles by the tumor and to determine the efficacy of the treatment.

1 Background and motivations

Breast cancer is the most commonly occurring cancer in women. In the last decades, advances in medicine and in surgery have improved life expectancy of people affected by the disease. However, the high toxicity associated to the chemotherapy remains a big challenge in breast cancer treatment. To overcome this issue, nanoparticles conjugated with cancer cell specific antibodies have been developed to improve drug delivery to tumors, while sparing healthy tissues. Mathematical modeling provides tools that are useful to better understand the properties of these nanoparticles.

The project is done in collaboration with the SMARTc team of the University of Marseille, who developed antibody nano-conjugates (ANCs) [1], that are currently evaluated in a preclinical phase. These nanoparticles consist in docetaxel-encapsulated liposomes engrafted with trastuzumab on the surface. The latter is an antibody that targets the Her2 receptors on the breast cancer cells. However, the intra-tumor penetration of this nanoparticles is not quantitatively known. The goal of this project is to use mathematical modeling to understand the dynamic of ANCs penetration in the tumor tissue and to determine their efficacy. In particular, the questions we want to address are the following: how are they transported in tumor tissues? Do they have the same efficacy of the other treatments? Is it possible to personalize the dose according to the characteristics of the lesion of a patient?

Biophysical models allow to understand the transport process of nanoparticles inside the tumor tissue. This is the first step to determine the efficacy of the ANCs.

2 Nanoparticle distribution in tumor tissues

When nanoparticles are intravenously injected, they have to overcome several barriers. First, they are transported to the tumor site through the vessels. Thanks to the leaky capillaries that characterize tumors, they are extravasated into the interstitial compartment. Finally, they are transported inside the interstitium and uptaken by the cells by endocytosis. Diffusion and convection are the main responsible of nanoparticle distribution in the tumor tissue [2]. Mathematical modeling allows to quantify the portion of drug that reaches the

tumor tissue. This is of fundamental importance to understand and improve the nanoparticle efficacy. Indeed, the therapy has to reach the target cells in an optimal quantity to be successful.

We focused on the development of a spatio-temporal model to describe the nanoparticle distribution inside the tumor tissue, in particular their penetration into the interstitial compartment. We defined a multiple scale model for nanoparticle transport in the tumor interstitium and capillaries. The tumor was considered as a porous medium irrigated by capillaries. Using the homogenization method [3, 4], we derived a macro-scale model of ANC transport. This approach allowed us to take into account the micro-structure at the macroscopic scale avoiding to solve a multiple scale model, that would have high computational costs. The derived model couples diffusion-convection-reaction equations of the nanoparticles in the tumor vessels and interstitium. The reaction term took into account the ANC uptake by the cells. Simulations of the model were obtained using the Finite Element Method. Qualitative results proved that the model is in agreement with biological observations. In particular, microscopic parameters such as the vessel geometry, permeability and diffusion coefficients play a significant role in the dynamic of penetration of ANCs in the tumor tissue. We plan to calibrate the model with experimental data provided by the SMARTc team. In particular, *in vitro* data quantify the efficacy of the treatment according to the nanoparticle concentration; *ex vivo* imaging data specify the microstructure of the tumor and the spatial distribution of the nanoparticles in the tumor tissue. An example of microvasculature of a tumor is provided in Fig 1.

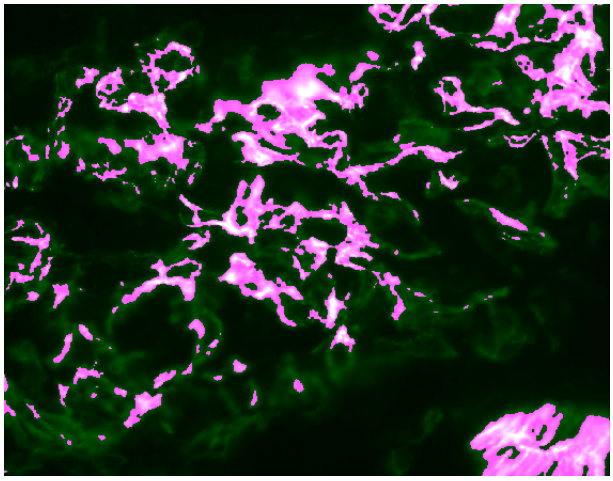

Figure 1: *Ex-vivo microscopy showing the tumor microvasculature.*

3 Efficacy of nanoparticle treatment

Experiments *in vivo* allowed to determine the efficacy of the ANCs and to compare it to standard therapies, such as chemotherapy and trastuzumab. Using mathematical models of tumor growth based on differential equations we can quantify the efficacy of the nanoparticle treatment on mice. Data fit to the models can be carried out within the statistical framework of nonlinear mixed effects modeling, that allows to take into account simultaneously the tumor growth dynamic and the inter-individual variability.

4 Future directions

We plan to calibrate the model with experimental data. This is of fundamental importance in order to quantify the ANC distribution in the tumor.
Eventually, we aim at understanding the link between the nanoparticle penetration into the tumor tissue and their efficacy in order to optimize the treatment with ANCs according to the tumor pathophysiology. This project has a direct impact on the development of the ANCs, that will be later tested in a clinical phase I.

Acknowledgments

This research has been partly funded by the Plan Cancer NUMEP (PC201615) and the Plan Cancer QUANTIC (19CM148-00). We acknowledge Anne Rodallec, Guillaume Sicard and Joseph Ciccolini for providing the data and for the useful hints and discussions.

References

[1] A. Rodallec, G. Sicard, S. Giacometti, M. Carré, B. Pourroy, F. Bouquet, A. Savina, B. Lacarelle, J. Ciccolini, and R. Fanciullino, "From 3D spheroids to tumor bearing mice: Efficacy and distribution studies of trastuzumab-docetaxel immunoliposome in breast cancer," *Int J Nanomedicine*, vol. Volume 13, pp. 6677–6688, Oct. 2018.

[2] R. K. Jain and T. Stylianopoulos, "Delivering nanomedicine to solid tumors," *Nature Reviews Clinical Oncology*, vol. 7, pp. 653–664, Nov. 2010.

[3] G. Allaire, "Homogenization and Two-Scale Convergence," *SIAM J. Math. Anal.*, vol. 23, pp. 1482–1518, Nov. 1992.

[4] R. J. Shipley and S. J. Chapman, "Multiscale Modelling of Fluid and Drug Transport in Vascular Tumours," *Bulletin of Mathematical Biology*, vol. 72, pp. 1464–1491, Aug. 2010.

Cohomology, (φ,τ)-modules and p-divisible groups

Luming Zhao
Directeur : Olivier Brinon

Laboratoire Bordelais de Recherche en Informatique, CNRS UMR 5800, Université de Bordeaux

Herr complex computes the Galois cohomology of a p-adic representation using its associated (φ,Γ)-module. We construct a similar complex using Caruso's (φ,τ)-modules and apply it to representations arising from p-divisible groups.

Let p be an odd prime, and K a complete discrete valuation field of mixed characteristic $(0,p)$, with perfect residue field k. Fix an algeabraic closure $\overline{K}$ of K and let $\widehat{\overline{K}}$ be its completion. Let $\pi \in K$ be a uniformizer, $E \in W(k)[X]$ the associated Eisenstein polynomial. Let $\mathscr{G}_K = \mathsf{Gal}(\overline{K}/K)$ and $\chi\colon \mathscr{G}_K \to \mathbf{Z}_p^\times$ the cyclotomic character. Put $\mathcal{R} = \varprojlim_{x\mapsto x^p} \mathcal{O}_{\widehat{\overline{K}}}$. Fix $\tilde{\pi} := (\pi_n)_{n\geq 0} \in \mathcal{R}$ with $\pi_0 = \pi$, and $\varepsilon := (\zeta_{p^i})_{i\geq 0} \in \mathcal{R}$ such that ζ_{p^i} is a primitive p^i-th root of unity. Put $K_\pi := \cup_n K(\pi_n)$, $\mathscr{G}_{K_\pi} := \mathsf{Gal}(\overline{K}/K_\pi)$, and $K_\zeta := \cup_n K(\zeta_{p^n})$ with $\mathscr{G}_{K_\zeta} := \mathsf{Gal}(\overline{K}/K_\zeta)$. Let $\Gamma := \mathsf{Gal}(K_\zeta/K)$ and $L = K_\pi K_\zeta$ with Galois group $H_\infty = \mathsf{Gal}(\overline{K}/L)$.

1 Breuil-Kisin modules

Endow $\mathfrak{S} := W(k)[\![u]\!]$ with the Frobenius endomorphism σ defined by $\sigma(u) = u^p$. A φ-module (over $\mathfrak{S}$) is an $\mathfrak{S}$-module $\mathfrak{M}$ equipped with a σ-semilinear map $\varphi : \mathfrak{M} \to \mathfrak{M}$. Let $\mathbf{Mod}^{1,fr}_{/\mathfrak{S}}$ denote the category of φ-modules of E-height 1, *i.e.* $\mathfrak{M}$ is finite free over $\mathfrak{S}$ and the cokernel of the linearization $\varphi^* = 1\otimes\varphi\colon \mathfrak{S}\otimes_{\sigma,\mathfrak{S}} \mathfrak{M} \to \mathfrak{M}$ is killed by E. Objects of $\mathbf{Mod}^{1,fr}_{/\mathfrak{S}}$ are also called Breuil-Kisin modules (or Kisin modules of height 1). Let $\mathbf{BT}_{/\mathcal{O}_K}$ denote the category of p-divisible groups over $\mathcal{O}_K$.

Theorem 5 (Kisin). *[1] There exists an equivalence of categories*

$$\mathbf{Mod}^{1,fr}_{/\mathfrak{S}} \simeq \mathbf{BT}_{/\mathcal{O}_K}.$$

The map $\mathfrak{S} = W(k)[\![u]\!] \hookrightarrow W(\mathcal{R})$, $\sum_{n\geq 0} a_n u^n \mapsto \sum_{n\geq 0} a_n[\tilde{\pi}]^n$ induces a $\mathscr{G}_{K_\pi}$-equivariant and φ-compatible embedding $j\colon \mathscr{O}_\mathscr{E} \hookrightarrow W(\mathsf{Frac}\mathcal{R})$, where $\mathscr{O}_\mathscr{E}$ denotes the p-adic completion of $\mathfrak{S}[u^{-1}]$. Let $\mathscr{O}_\mathscr{E}^{nr}/\mathscr{O}_\mathscr{E}$ be the maximal unramified extension of $\mathscr{O}_\mathscr{E}$ and $\widehat{\mathscr{O}_\mathscr{E}^{nr}}$ its p-adic completion. These maps are used to relate (generalizations of) Breuil-Kisin modules to p-adic Hodge theory.

2 (φ,Γ)-modules

Replacing $[\tilde{\pi}]$ by $[\varepsilon]-1$ in the preceding construction, we construct an analogue $\mathbf{A}$ of $\widehat{\mathscr{O}_\mathscr{E}^{nr}}$ in $W(\mathsf{Frac}\mathcal{R})$. In contrast with $\widehat{\mathscr{O}_\mathscr{E}^{nr}}$, $\mathbf{A}$ has an action of $\mathscr{G}_K$.

Put $\mathbf{A}_K := \mathbf{A}^{\mathscr{G}_{K_\zeta}}$, which has a linear, continuous action of Γ and a continuous Frobenius φ commuting with the action of Γ. Define the functor

$$D : T \mapsto D(T) = (\mathbf{A}\otimes_{\mathbf{Z}_p} T)^{\mathscr{G}_{K_\zeta}}$$

from $\mathbf{Rep}_{\mathbf{Z}_p}(\mathscr{G}_K)$, the category of $\mathbf{Z}_p$-adic representations of $\mathscr{G}_K$, to the category $\mathbf{Mod}^{\text{ét}}_{\mathbf{A}_K}(\varphi,\Gamma)$ of étale (φ,Γ)-modules on $\mathbf{A}_K$ whose objects are finitely generated $\mathbf{A}_K$-modules with commuting semi-linear actions of Γ and φ, where φ is étale. Define also

$$V : M \mapsto V(M) = (\mathbf{A}\otimes_{\mathbf{A}_K} M)^{\varphi=1}$$

from the category $\mathbf{Mod}^{\text{ét}}_{\mathbf{A}_K}(\varphi,\Gamma)$ to $\mathbf{Rep}_{\mathbf{Z}_p}(\mathscr{G}_K)$.

Theorem 6 (Fontaine). *[2] The following natural maps are isomorphisms*

$$\mathbf{A}\otimes_{\mathbf{A}_K} D(T) \to \mathbf{A}\otimes_{\mathbf{Z}_p} T$$

$$\mathbf{A}\otimes_{\mathbf{Z}_p} V(M) \to \mathbf{A}\otimes_{\mathbf{A}_K} M.$$

In particular, D and V are quasi-inverse equivalences of categories between $\mathbf{Rep}_{\mathbf{Z}_p}(\mathscr{G}_K)$ and $\mathbf{Mod}^{\text{ét}}_{\mathbf{A}_K}(\varphi,\Gamma)$.

One can easily compute the cohomology of a p-adic representation in terms of its (φ,Γ)-module.

Theorem 7 (Herr). *[3] Fix γ a topological generator of Γ, then the cohomology of the complex*

$$0 \to D(T) \xrightarrow{f_1} D(T)\oplus D(T) \xrightarrow{f_2} D(T) \to 0$$

(where maps $f_1 = \begin{pmatrix}\varphi-1\\ \gamma-1\end{pmatrix}$ and $f_2 = (\gamma-1, 1-\varphi)$) identifies canonically and functorially with the Galois cohomology of the representation T.

This important result has many applications in p-adic Hodge theory, Iwasawa theory, etc.

3 (φ, τ)-modules

It is difficult to relate Breuil-Kisin modules and their generalizations with (φ, Γ)-modules, because the extensions K_ζ and K_π are linearly disjoint, and K_π/K is not Galois. The good objects in this context to consider are (φ, τ)-modules introduced by Caruso.

Notation 1. *Let $F_\tau = \mathsf{Frac}\mathcal{R}^{H_\infty}$, it obviously contains $F_0 = k(\!(\tilde{\pi})\!) \subset \mathsf{Frac}\mathcal{R}$. Fix $\tau \in \mathsf{Gal}(\overline{K}/K_\zeta)$ such that $\tau(\tilde{\pi}) = \varepsilon\tilde{\pi}$. We now construct a ring*

$$\mathcal{E}^{\mathsf{int}} = \{\sum_{i\in\mathbf{Z}} a_i u^i;\ a_i \in W(k),\ \lim_{i\to-\infty} a_i = 0\}$$

and embed it into $W(\mathsf{Frac}\mathcal{R})$ by sending u to $[\tilde{\pi}]$. Endow $\mathcal{E}^{\mathsf{int}}$ with the p-adic valuation: it is a discrete valuation ring, with residue field F_0. Let $\mathcal{E}^{ur,\mathsf{int}}$ be the unique ind-étale sub-algebra of $W(\mathsf{Frac}\mathcal{R})$ whose residue field is $F_0^{sep} \subset \mathsf{Frac}\mathcal{R}$ and $\widehat{\mathcal{E}^{ur,\mathsf{int}}}$ its p-adic completion. We put $\mathcal{E}_\tau^{\mathsf{int}} = W(F_\tau)$ and $\mathcal{E}_\tau = \mathsf{Frac}\mathcal{E}_\tau^{\mathsf{int}}$.

Definition 1. *A (φ, τ)-module over $(\mathcal{E}^{\mathsf{int}}, \mathcal{E}_\tau)$ consists of*

(i) an étale φ-module M over $\mathcal{E}^{\mathsf{int}}$;

(ii) a τ-semi-linear endomorphism τ_M on $M_\tau := \mathcal{E}_\tau^{\mathsf{int}} \otimes_{\mathcal{E}^{\mathsf{int}}} M$ which commutes with $\varphi_{\mathcal{E}_\tau^{\mathsf{int}}} \otimes \varphi_M$ and which satisfies, for all $g \in G_\infty/H_\infty$ such that $\chi(g) \in \mathbf{N}$, the following relation:

$$(\forall x \in M)\ (g \otimes \mathsf{id}) \circ \tau_M(x) = \tau_M^{\chi(g)}(x)$$

Let $\mathbf{Mod}_{\mathcal{E}^{\mathsf{int}}, \mathcal{E}_\tau^{\mathsf{int}}}(\varphi, \tau)$ be the corresponding category.

Theorem 8. *[4][§1.3] We define two functors $\mathcal{D}$ and $\mathcal{C}$, which give quasi-inverse categorical equivalences as follows:*

$$\mathbf{Rep}_{\mathbf{Z}_p}(\mathscr{G}_K) \simeq \mathbf{Mod}_{\mathcal{E}^{\mathsf{int}}, \mathcal{E}_\tau^{\mathsf{int}}}(\varphi, \tau)$$
$$T \mapsto \mathcal{D}(T) = (\widehat{\mathcal{E}^{ur,\mathsf{int}}} \otimes T)^{G_\infty}$$
$$\mathcal{C}(D) = (\widehat{\mathcal{E}^{ur,\mathsf{int}}} \otimes_{\mathcal{E}^{\mathsf{int}}} D)^{\varphi=1} \hookleftarrow D$$

Notation 2. *If $D \in \mathbf{Mod}_{\mathcal{E}^{\mathsf{int}}, \mathcal{E}_\tau^{\mathsf{int}}}(\varphi, \tau)$, put $D_\tau := \mathcal{E}_\tau^{\mathsf{int}} \otimes_{\mathcal{E}^{\mathsf{int}}} D$ and $D_{\tau,0} := \{x \in D_\tau;\ (\forall g \in G_\infty)\ \chi(g) \in \mathbf{N} \Rightarrow (g \otimes 1)(x) = x + \tau_D(x) + \cdots + \tau_D^{\chi(g)-1}(x)\}$.*

Definition 2. *Let $D \in \mathbf{Mod}_{\mathcal{E}^{\mathsf{int}}, \mathcal{E}_\tau^{\mathsf{int}}}(\varphi, \tau)$. We define a complex $\mathcal{C}_{\varphi,\tau}(D)$ as follows:*

$$\begin{array}{rccc} 0 \to D \to & D \bigoplus D_{\tau,0} & \to D_{\tau,0} \to 0 \\ x \mapsto & (\rho(x), \gamma(x)) & \\ & (y, z) & \mapsto \gamma(y) - \rho(z) \end{array}$$

where the first term is of degree -1.

Theorem 9. *For all natural integers i, there is a canonical and functorial isomorphism*

$$H^i(\mathcal{C}_{\varphi,\tau}(\mathcal{D}(T)) \simeq H^i(\mathscr{G}_K, T).$$

Remark 1.

(1) This is already observed by Caruso when T is killed by p and $i \in \{0, 1\}$.

(2) Our complex is a refinement of the four terms complex constructed by Tavares Ribeiro. [5, Theorem 1.5]

Extending the scalars to the Robba ring, there is a complex similar to $\mathcal{C}_{\varphi,\tau}$ involving the logarithm of τ, that should compute the cohomology of an open subgroup of $\mathscr{G}_K$.

As an application of what precedes, we plan to use theorem 9 to provide a three terms complex that computes the Galois cohomology of the Tate module of a p-divisible group in terms of its Breuil-Kisin module.

References

[1] Kisin, Mark, *Crystalline representations and F-crystals*, Algebraic geometry and number theory, Progress in Mathematics 253, Birkhäuser Boston, (2006)

[2] Fontaine, Jean-Marc, *Représentations p-adiques des corps locaux (1ère)*, The Grothendieck Festschrift, Birkhäuser Boston, (1991)

[3] Herr, Laurent, *Sur la cohomologie galoisienne des corps p-adiques*, Bulletin de la Société mathématique de France 126, (1998)

[4] Caruso, Xavier, *Representations galoisiennes p-adiques et (ϕ, τ)-modules*, Duke Math, (2013)

[5] Tavares Ribeiro, Floric *An explicit formula for the Hilbert symbol of a formal group* Annales de l'Institut Fourier 61, (2011)

L'Ecole doctorale de mathématiques de Bordeaux a été créée en 1991 et élargie à l'informatique en 1995. Elle a été dirigée successivement par :

Henri Cohen	:	1995-1998
Yves Métivier	:	1999-2002
Thierry Colin	:	2003-2005
Éric Sopena	:	2006-2009
Christophe Bavard	:	2010-2013
Nicolas Hanusse	:	2014-2017
Andreas Hartmann	:	2018-