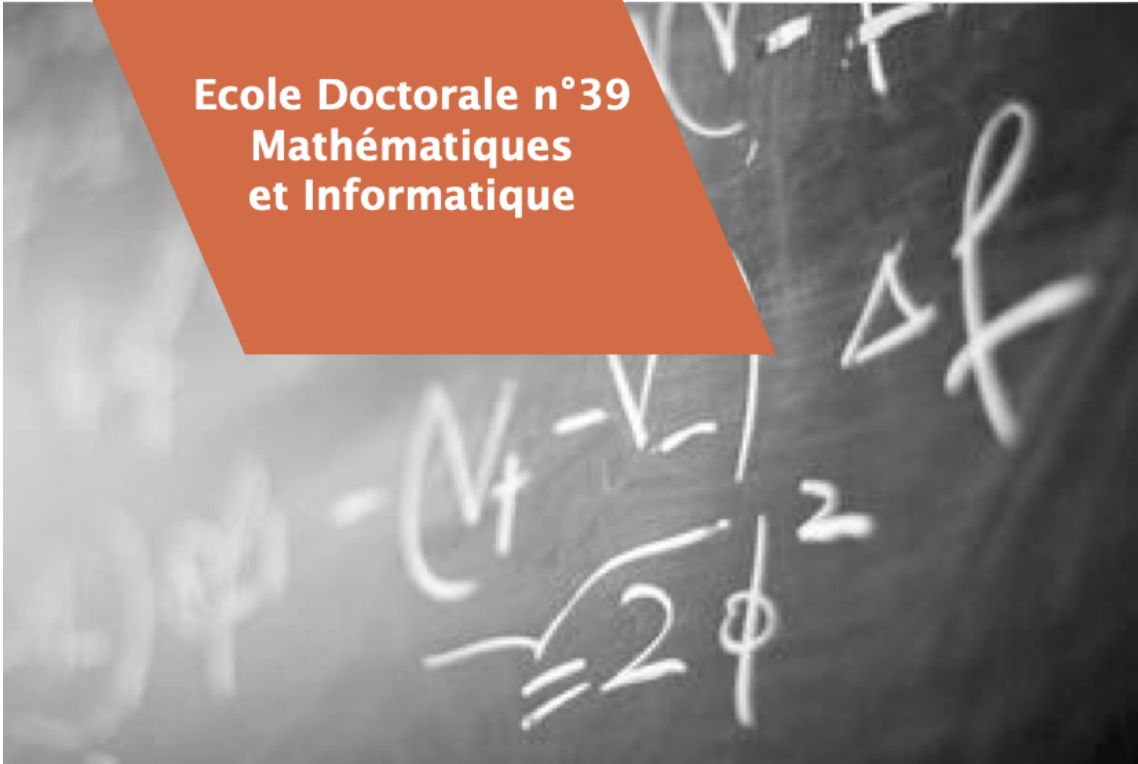




**Ecole Doctorale n°39  
Mathématiques  
et Informatique**

**Résumés Étendus des Thèses de Doctorat  
de Troisième Année 2023**

**École 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 Turing)*

# École Doctorale Mathématiques et Informatique ED 39

Université de Bordeaux

Année 2024

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 59 résumés étendus : 33 pour l'informatique et 26 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'Ayres, cardinal 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'Algorithmes 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.

### **Direction 2022**

- Guillaume Blin (directeur)
- Nicolas Bonichon (directeur adjoint - formation informatique)
- Sylvain Ervedoza (directeur adjoint - formation mathématique)
- Isabelle Bordas (gestionnaire)

Site internet de l'ED : <https://ed-mi.u-bordeaux.fr/>



# Table des matières

|                                                                                                                                         |          |
|-----------------------------------------------------------------------------------------------------------------------------------------|----------|
| <b>I Informatique</b>                                                                                                                   | <b>7</b> |
| 1 Caractérisation et suivi des charges de travail HPC de nouvelle génération<br><i>Alexis BANDET</i>                                    | 9        |
| 2 Optimizing GAN for data synthesis : application to medical imaging and affective computing<br><i>Zakariae BELMEKKI</i>                | 11       |
| 3 Language and RL a beneficial interaction<br><i>Thomas CARTA</i>                                                                       | 13       |
| 4 Exploring the impact of neural network architecture on decision-making processes<br><i>Naomi CHAIX-EICHEL</i>                         | 15       |
| 5 Identifier robustement des biomarqueurs de cancers et explorer leur intérêt clinique<br><i>Elsa CLAUDE</i>                            | 17       |
| 6 Reachability in temporal graphs<br><i>Timothée CORSINI</i>                                                                            | 19       |
| 7 Développement d'infrastructure numérique pour la sûreté ferroviaire<br><i>Mayssa DARDOUR</i>                                          | 21       |
| 8 StarONNX : a Dynamic Scheduler for Low Latency and High Throughput Inference on Heterogeneous Resources<br><i>Jean-François DAVID</i> | 23       |
| 9 Analyse automatisée de binaires à la recherche de vulnérabilités matérielles<br><i>Theo DE CASTRO PINTO</i>                           | 25       |
| 10 Computer-assisted rough 2D animation inbetweening<br><i>Melvin EVEN</i>                                                              | 27       |
| 11 Modeling the appearance of material iridescence based on periodic structures<br><i>Gary FOURNEAU</i>                                 | 29       |
| 12 Bioinformatic methods to study metabolism and cell types in Glioblastoma<br><i>Johanna GALVIS</i>                                    | 31       |
| 13 Exploring virtual reality by and for non-expert users<br><i>Edwige CHAUVERGNE</i>                                                    | 33       |
| 14 Modeling the eco-evolutionary origins of open-ended skill acquisition<br><i>Gautier HAMON</i>                                        | 35       |
| 15 Algorithmes De Couverture Pour Les Réseaux De Petri Certifiés En Coq<br><i>Thibault HILAIRE</i>                                      | 37       |
| 16 Practical white-box cryptography<br><i>Agathe HOUZELOT</i>                                                                           | 39       |

|                                                                                                                                                                                                     |           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>17 Chasse aux mines sous-marines à l'aide d'apprentissage par renforcement multi-agents</b><br><i>Adrien JÉGOUREL</i>                                                                            | <b>41</b> |
| <b>18 Towards socio-cultural artificial agents</b><br><i>Grgur KOVAČ</i>                                                                                                                            | <b>43</b> |
| <b>19 Méthodes de réduction d'ordre-partiel pour la vérification de systèmes concurrents temps-réel</b><br><i>Sarah LARROZE</i>                                                                     | <b>45</b> |
| <b>20 Les pratiques de code : de la documentation à la détection</b><br><i>Corentin LATAPPY</i>                                                                                                     | <b>47</b> |
| <b>21 Reconfiguration and Kempe changes</b><br><i>Clément LEGRAND-DUCHESNE</i>                                                                                                                      | <b>49</b> |
| <b>22 Advancing Lung Imaging : Deep Learning for MR-to-CT Synthesis</b><br><i>Arthur LONGUEFOSSE</i>                                                                                                | <b>51</b> |
| <b>23 Modeling Apperance of Porous Materials</b><br><i>Simon LUCAS</i>                                                                                                                              | <b>53</b> |
| <b>24 Interactions Tangibles pour l'Exploitation Ferrovaire dans les Centres Opérationnels</b><br><i>Maudeline MARLIER</i>                                                                          | <b>55</b> |
| <b>25 Vérification et synthèse de systèmes distribués à synchronisation faible</b><br>ou comment appliquer la théorie des automates à la vérification de systèmes distribués<br><i>Corto MASCLE</i> | <b>57</b> |
| <b>26 Program Synthesis : Scaling Enumerative Approaches</b><br><i>Théo MATRICON</i>                                                                                                                | <b>59</b> |
| <b>27 Using sequence analysis to characterize behaviors during a creative problem solving task</b><br><i>Axel PALAUDE</i>                                                                           | <b>61</b> |
| <b>28 Optimization of task-based applications on heterogeneous computing nodes</b><br><i>Hayfa TAYEB</i>                                                                                            | <b>63</b> |
| <b>29 Improving motor rehabilitation after a stroke with artificial intelligence and non-invasive neurotech-<br/>nologies</b><br><i>David TROCELLIER</i>                                            | <b>65</b> |
| <b>30 Rendu Expressif et Contrôle de Lignes</b><br><i>Panagiotis TSIAPKOLIS</i>                                                                                                                     | <b>67</b> |
| <b>31 EEG-based Decoding of Aesthetic Preference for Visual Art</b><br><i>Marc WELTER</i>                                                                                                           | <b>69</b> |
| <b>32 Cylinders in square-tiled surfaces and metric ribbon graphs</b><br><i>Ivan YAKOVLEV</i>                                                                                                       | <b>71</b> |
| <b>33 Rockmate : memory efficient training for neural networks</b><br><i>Xunyi ZHAO</i>                                                                                                             | <b>73</b> |
| <b>II Mathématiques</b>                                                                                                                                                                             | <b>75</b> |
| <b>34 Modeling and decomposition methods for robust facility location and network design problems<br/>with integer recourse</b><br><i>Komlanvi Parfait AMETANA</i>                                  | <b>77</b> |
| <b>35 Condensed duality theorems for the Weil group of a <math>p</math>-adic field</b><br><i>Marco ARTUSA</i>                                                                                       | <b>79</b> |

|                                                                                                                                         |            |
|-----------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>36 Cryptosystems based on class groups of imaginary quadratic fields</b>                                                             | <b>81</b>  |
| <i>Agathe BEAUGRAND</i>                                                                                                                 |            |
| <b>37 Non-convex methods for inverse problems with low dimensional models</b>                                                           | <b>83</b>  |
| <i>Pierre-Jean BÉNARD</i>                                                                                                               |            |
| <b>38 Couplages de mouvements browniens en géométrie sous-riemannienne</b>                                                              | <b>85</b>  |
| <i>Magalie BÉNÉFICE</i>                                                                                                                 |            |
| <b>39 Analyse et Développement d'algorithmes d'optimisation pour les réseaux de neurones</b>                                            | <b>87</b>  |
| <i>Bilel BENSAID</i>                                                                                                                    |            |
| <b>40 Radiomic Analysis for relapse prediction in diffuse large B-cell lymphoma</b>                                                     | <b>89</b>  |
| <i>Khaoula CHAHDI</i>                                                                                                                   |            |
| <b>41 Finite volume methods for the cell-by-cell model of cardiac electrophysiology</b>                                                 | <b>91</b>  |
| <i>Zeina CHEHADE</i>                                                                                                                    |            |
| <b>42 Toward robust linear implicit schemes for steady states supersonic flows.</b>                                                     | <b>93</b>  |
| <i>Benoit COSSART</i>                                                                                                                   |            |
| <b>43 Statistical learning through the scope of random matrix theory and free probability</b>                                           | <b>95</b>  |
| <i>Issa-Mbenard DABO</i>                                                                                                                |            |
| <b>44 mDAE : modified Denoising AutoEncoder for missing data imputation</b>                                                             | <b>97</b>  |
| <i>Mariette DUPUY</i>                                                                                                                   |            |
| <b>45 Parametrization free structure-texture decomposition</b>                                                                          | <b>99</b>  |
| <i>Antoine GUENNEC</i>                                                                                                                  |            |
| <b>46 L'influence de la structure du groupe de Galois sur la répartition des automorphismes de Frobenius</b>                            | <b>101</b> |
| <i>Mounir HAYANI</i>                                                                                                                    |            |
| <b>47 Computer vision for Electroporation-based ablation of deep-seated tumour</b>                                                      | <b>103</b> |
| <i>Eloïse INACIO</i>                                                                                                                    |            |
| <b>48 Modelling of aero-elasticity in large transformations by a partitioned coupling : application to large wind turbines.</b>         | <b>105</b> |
| <i>Caroline LE GUERN</i>                                                                                                                |            |
| <b>49 Artificial Intelligence-assisted tool for HEP-2 Antinuclear Antibody Patterns' recognition</b>                                    | <b>107</b> |
| <i>Guillaume MARTINROCHE</i>                                                                                                            |            |
| <b>50 Amélioration du suivi clinique de tumeurs intracrâniennes à l'aide d'équations différentielles et de méthodes d'apprentissage</b> | <b>109</b> |
| <i>Virginie MONTALIBET</i>                                                                                                              |            |
| <b>51 Optimal stochastic control and application on passive trajectography</b>                                                          | <b>111</b> |
| <i>Romain NAMYST</i>                                                                                                                    |            |
| <b>52 Electrocardiology Modeling after Catheter Ablations for Atrial Fibrillation</b>                                                   | <b>113</b> |
| <i>Simone NATI POLTRI</i>                                                                                                               |            |
| <b>53 Simulations numériques standardisées de dispositifs de stimulation électrique cardiaque</b>                                       | <b>115</b> |
| <i>Valentin PANNETIER</i>                                                                                                               |            |
| <b>54 Lois de paroi à apprentissage profond pour simulations aérodynamiques</b>                                                         | <b>117</b> |
| <i>Michele ROMANELLI</i>                                                                                                                |            |
| <b>55 The Littlewood problem and non harmonic Fourier series</b>                                                                        | <b>119</b> |
| <i>Chadi SABA</i>                                                                                                                       |            |

|           |                                                                                                                                             |            |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>56</b> | <b>Approches Galerkin Discontinues spatio-temporelles sur des mailles hiérarchiques pour les équations de Navier-Stokes incompressibles</b> | <b>121</b> |
|           | <i>Alexis TARDIEU</i>                                                                                                                       |            |
| <b>57</b> | <b>Effet régularisants, anisotropie et controlabilité en mécanique des fluides</b>                                                          | <b>123</b> |
|           | <i>Adrien TENDANI-SOLER</i>                                                                                                                 |            |
| <b>58</b> | <b>Multivariate quantiles, optimal transport</b>                                                                                            | <b>126</b> |
|           | <i>Gauthier THURIN</i>                                                                                                                      |            |
| <b>59</b> | <b>Méthodes numériques cinétiques novatrices pour la radiothérapie et les plasmas</b>                                                       | <b>128</b> |
|           | <i>Thomas VIGIER</i>                                                                                                                        |            |



**Première partie**

**Informatique**



# CARACTÉRISATION ET SUIVI DES CHARGES DE TRAVAIL HPC DE NOUVELLE GÉNÉRATION

Alexis BANDET

Direction : Francieli BOITO

Co-Direction : Guillaume PALLEZ

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

Le calcul haute performance (HPC) est à l'avant-garde des défis les plus importants de la société, de la lutte contre le coronavirus et la recherche de nouveaux traitements médicamenteux, à la modélisation et la prévention des impacts du réchauffement climatique. Dans le contexte de la convergence entre le HPC et le big data, la notion de tâche scientifique évolue vers celle de flux de travail. Ces flux se composent de tâches à forte intensité de calcul et de données. Les ressources HPC sont arbitrées entre applications soumises en utilisant très peu d'information : généralement le nombre de processeurs ainsi que le temps d'exécution estimé. Cependant, bien que s'exécutant sur des nœuds indépendants, les applications concurrentes partagent l'infrastructure d'E/S. Ceci signifie que les interférences avec d'autres applications peuvent nuire à leurs performances, surtout si elles sont gourmandes en données. En outre, les interférences ralentissent les applications de manière non uniforme, car le facteur de ralentissement de chacune d'entre elles dépend fortement des caractéristiques d'accès de l'ensemble. Ces problèmes devraient s'aggraver à mesure que les nouvelles charges de travail HPC incluent des codes plus diversifiés. Dans cette thèse, nous envisageons d'étudier des stratégies de suivi minimales et efficaces pour les techniques d'arbitrage de ressources. Pour ce faire, nous prévoyons de commencer par étudier la caractérisation possible de classes d'applications pour concevoir des modèles généraux. Nous étudierons le volume minimal d'information qu'un outil de suivi devrait savoir mesurer pour chaque nouveau job pour être capable de donner une bonne approximation de sa classe. Ces informations doivent également permettre qu'un algorithme d'ordonnancement les utilise tout en restant assez efficace. Nous travaillerons ensuite à l'élaboration d'un outil de suivi permettant d'obtenir ces données.

## Allocation dynamique de ressources

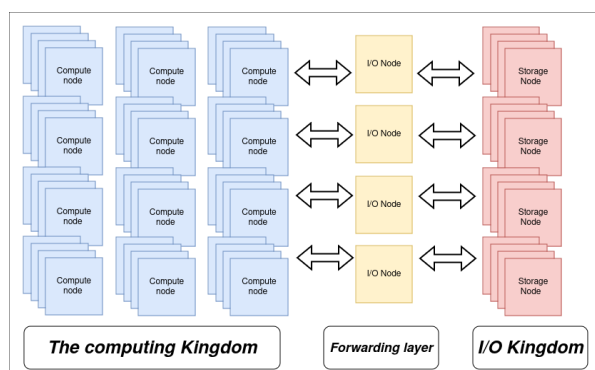


FIGURE 1.1 – Diagramme simplifié d'une machine HPC.

La première partie de la thèse s'est concentrée sur la compréhension et l'investigation de la distribution des ressources de stockage des données persistantes dans un supercalculateur [1]. Dans un tel environnement, ces ressources, telles que les Object Storage Target (OST) ou les I/O Forwarding Nodes, sont partagées entre les applications. Chaque ressource peut alors être sollicitée par plusieurs applications simultanément, ce qui crée de la congestion. Les applications synchrones, dont le comportement est décrit par la figure 1.2 sont alors

ralenties.

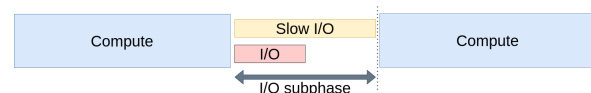


FIGURE 1.2 – Modèle d'application synchrone avec 2 ressources d'E/S.

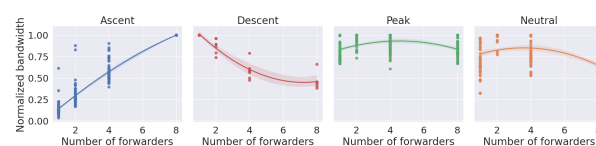


FIGURE 1.3 – Profils d'E/S des applications, catégorisées.

Mais si les systèmes de stockage sont parallèles, c'est avant tout pour pouvoir multiplier la bande passante de lecture et d'écriture. Pour une application avoir accès simultané à plus de ressources n'est pas toujours corrélé à plus de performances. La figure 1.3 montre que si certaines applications profitent d'une plus grande parallélisation des ressources, sur d'autres cela n'a pas d'impact, voire un impact négatif.

Ces observations nous permettent de dessiner le contour de deux problèmes. Le premier est de trouver le nombre de ressources à attribuer à chaque application. Un nombre insuffisant de ressources pénalisera les applica-

tions qui n'atteindront pas leur performance maximale, alors que trop de ressources va augmenter la demande sur le système, et donc créer de la congestion. Le second problème est de savoir comment répartir les applications sur le système afin de minimiser les collisions (quand deux applications accèdent simultanément à la même ressource) et répartir équitablement la charge sur le système. Nous proposons alors deux catégories d'algorithmes : les allocator et les placement. Nous comparons plusieurs algorithmes de chaque catégorie.

Se pose alors la question de la quantité de données nécessaires pour chaque algorithmes. En effet, chaque algorithmes va nécessiter un certain nombre d'entrée qui vont être plus ou moins difficile à obtenir. Nous comparons les différents algorithmes sur les performances mais aussi sur leur coût de mise en oeuvre. Nous montrons aussi que même avec des informations incomplètes ou partielles certains algorithmes arrivent à avoir de très bonne performances, avec *in fine* un cout de déploiement réduit.

## Caractérisation par fréquence

Se pose ensuite la question de l'obtention des informations nécessaires, avec en premier lieu la fréquence des applications. Pouvoir prédire quand arrivera la prochaine phase d'I/O permet à des algorithmes de placement d'anticiper les besoins et donc d'éviter *a priori* les collisions. De plus, cette information peut bénéficier à d'autre système tel que les Burst Buffer ou d'autre système de placement des applications ou des E/S. Nous avons donc développé FTIO, une bibliothèque permettant d'obtenir en temps réel les caractéristiques temporelles d'une applications, et plus particulièrement sa fréquence (alternation entre phase de calculs intensifs et phases de lecture/écriture de données).

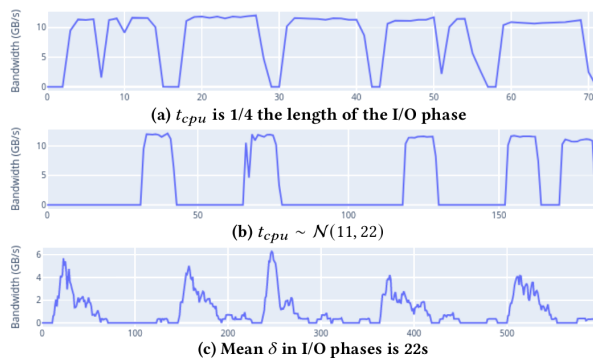


FIGURE 1.4 – Exemples d'applications semi-synthétiques.

Pour cela nous nous basons sur la transformée de Fourier afin d'extraire la fréquence principale de notre applications, et donc de connaître sa période. Nous fournissons une étude de robustesse de notre approche au moyen d'application semi-synthétique, telle que montrées dans figure 1.4. Cette approche permet de très facilement changer le niveau de bruit présent dans nos applications.

## Model de prédiction du nombre de ressource idéal

La seconde information essentiel au placement et allocation des applications est le profils I/O des applications, tel qu'illustré par la figure 1.3. Mais comme démontré plus tôt, la précision n'est pas vital pour les prédiction. Une prédiction de la bande passante n'est même pas un élément essentiel pour nos algorithmes. Nous introduisons donc un nouveau model de prédiction du nombre de ressources permettant à une applications d'atteindre son pic de performance E/S ainsi que la catégorie de l'application. Nous utilisons des paramètres intrinsèques à l'applications, comme la quantité de donnée, le schéma d'accès aux données ou la spatialité de ces dernières. Ce travail permet aussi de mettre en lumière les paramètres des applications qui ont une influence sur le profile ou la capacité de l'application à performer ou non en fonction d'un nombre de noeud.

## Bibliographie

- [1] Bandet, A., Boito, F., Pallez, G. (2024). Scheduling distributed I/O resources in HPC systems.
- [2] Tarraf, A., Bandet, A., Boito, F., Pallez, G., Wolf, F. (2023). Ftio : Detecting i/o periodicity using frequency techniques. arXiv preprint arXiv :2306.08601.
- [3] Bez, J. L., Miranda, A., Nou, R., Boito, F. Z., Cortes, T., Navaux, P. (2021, May). Arbitration policies for on-demand user-level i/o forwarding on hpc platforms. In 2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (pp. 577-586). IEEE.

# OPTIMIZING GAN FOR DATA SYNTHESIS : APPLICATION TO MEDICAL IMAGING AND AFFECTIVE COMPUTING

Zakariae BELMEKKI

Direction : Nadine COUTURE

Co-Direction : Karl JENKINS

Encadrement : David GOMEZ, Patrick REUTER

*EstiaR, Univ. Bordeaux, INRIA, Cranfield University*

This document presents the progress of the author's doctoral project on Generative Adversarial Networks (GANs). The latter have emerged as a powerful class of generative models, achieving state-of-the-art levels of realism on multiple types of data (images, 3D data, time-series). However, GANs are still very hard to train and many models, despite their depth, fail to achieve high levels of realism without careful tuning and model building. Therefore, the objective of this PhD is to optimise GAN to improve its performance. Also, it will be applied to two fields, medical imaging and affective computing.

## Generative Adversarial Networks

Generative Adversarial Networks (GANs) [7] are a class of generative models that are commonly used in multiple domains. Their impressive generative capabilities has earned them a lot of attention from the research community and thus, in the past decade, there has been a plethora of works on GANs.

The GAN framework consists of two models, a generator and a discriminator, trained simultaneously in an adversarial manner. The generator is trained to generate fake samples that fool the discriminator into classifying them as real samples, and the discriminator is trained to correctly classify real and fake samples. The minimax game is formulated like the following

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (2.1)$$

with  $G$  the generator,  $D$  the discriminator,  $p_{data}$  the real data distribution and  $p_z$  a distribution from which latent noise is drawn.

## GAN for Medical Imaging

The first work, which constitutes a sort of preliminary dive into GANs and their optimization, was on using GAN in a Medical Imaging context. The baseline architecture was the Deep Convolutional GAN (DCGAN) [1], which was adopted for generating x-ray chest images from the [2] dataset. The main contribution of the work was to optimize the DCGAN architecture for better and faster GAN convergence. The main changes to the baseline architecture and configuration are the following :

1. The activation functions, Relu and LeakyReLU, were replaced by ELU [3] in both the generator and the discriminator. The latter was reported to improve deep learning model training and is also

a requirement for the Two Time-Scale Update Rule (TTUR) (see details below).

2. TTUR [4] was adopted for training the model. It consists of choosing different learning rates for the generator and the discriminator in a way that allows the latter to train enough while the former does not cause significant perturbations.
3. Most state-of-the-art GAN models use Adam for training. In our case, we deviate from the mainstream and use a new optimizer, AdaS (Adaptive Step-Size) [5], which led to faster and better convergence.

The aforementioned changes led to an increase in sample realism as well as faster model convergence. The readers can read more about our work and results in the original conference paper [6].

## Reducing Similarity in Convolutional Networks

As a second step in optimizing GAN, we adopt a more global perspective, which consists of improving Convolutional Neural Networks (CNNs) since most GAN models are convolutional. While delving into convolutional feature maps, it was observed that multiple feature maps have a high degree of similarity : they exhibit the same features and are highly correlated. Further experiments showed that the standard way of dealing with such redundancy, mainly by reducing correlation of convolutional filters, is not always effective and is subject to high fluctuations depending on the weights and input data. After a thorough examination of the underlying theory, we propose a new regularization term that effectively reduces feature map similarity.

Tested on some standard classifiers trained on CIFAR-10 [8], this new regularization method led to higher accuracy, especially when used prior to training. Some

of the results are presented in Figures 2.1 and 2.2.  $\alpha$  and  $\beta$  are two parameters that control the way our regularization term is used. The cases where  $\alpha \neq 0$  and  $\beta \neq 0$  correspond to the cases where the regularization method is used. Notice that in all cases, our method leads to significantly better accuracy. Note that, since the work is not published yet, only minimal information is presented in this document.

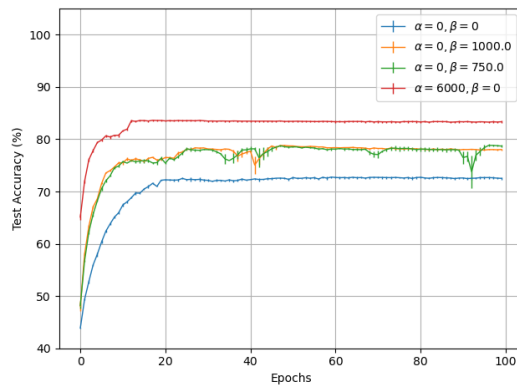


FIGURE 2.1 – ResNet18 trained with different settings. The cases where  $\alpha \neq 0$  and  $\beta \neq 0$

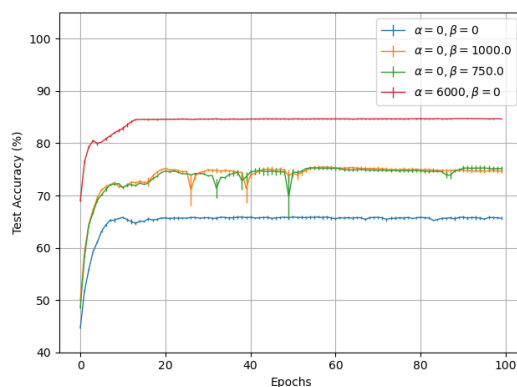


FIGURE 2.2 – Caption

The next step is to apply this method to a GAN to see whether it will improve its training or not.

## GAN for Affective Computing

In this work, GAN is used to generate sequence of emotional data in the form of action units. Since the data at hand are time-series, common practice dictates the use of Recurrent Neural Networks (RNN). However, we deviate from that mainstream choice and use a CNN, since the doctoral project focuses on convolutional GANs. The main contributions of this work, which is also yet to be published, are the following :

1. We propose a new configuration for CNNs, that is adapted to this kind of application based on theoretical concerns.
2. We adapt the Fréchet Inception Distance [4] to GANs for action unit sequences and thus provide a metric for the affective computing community to evaluate GANs for emotional sequential data.

3. We generate realistic controllable emotional data using an imbalanced and scarce dataset with multiple classes representing both simple and complex emotions.

## Bibliographie

- [1] Alec Radford and Luke Metz and Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2016
- [2] VinBigData Chest X-ray Abnormalities Detection, <https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection/overview>
- [3] Djork-Arné Clevert and Thomas Unterthiner and Sepp Hochreiter, Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), 2016
- [4] Martin Heusel and Hubert Ramsauer and Thomas Unterthiner and Bernhard Nessler and Sepp Hochreiter, GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, 2018
- [5] Yanai Eliyahu, AdasOptimizer, <https://github.com/YanaiEliyahu/AdasOptimizer>, 2013
- [6] Belmekki, Zakariae and Li, Jun and Jenkins, Karl and Reuter, Patrick and Gómez Jáuregui, David Antonio, 2022 7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS), An Empirical Evaluation of Generative Adversarial Nets in Synthesizing X-ray Chest Images, 2022
- [7] Ian J. Goodfellow and Jean Pouget-Abadie and Mehdi Mirza and Bing Xu and David Warde-Farley and Sherjil Ozair and Aaron Courville and Yoshua Bengio, Generative Adversarial Networks, 2014
- [8] Krizhevsky, Alex and Hinton, Geoffrey, Technical report, University of Toronto, Learning multiple layers of features from tiny images, 2009.

# LANGUAGE AND RL A BENEFICIAL INTERACTION

Thomas CARTA

Direction : Pierre-Yves OUDEYER<sup>1</sup>

Co-Direction : Olivier SIGAUD<sup>2</sup>

Encadrement : Sylvain LAMPRIER<sup>3</sup>

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

1 : Inria (Flowers), Université de Bordeaux, France

2 : Sorbonne Université, ISIR, Paris, France

3 : Université d'Angers, LERIA, SFR MATHSTIC, Angers, France

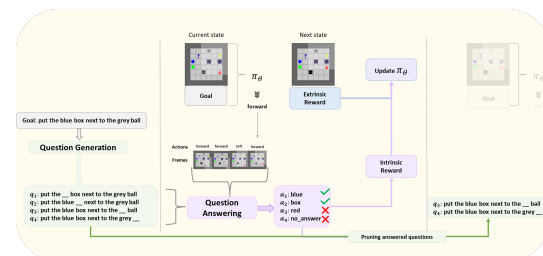
**W**e study how reinforcement learning (RL) and language can interact, in particular how language can benefit the exploration and learning of new tasks. We present two works that shed light on this relationship, the first showing how language can facilitate exploration and the second presenting a method for grounding a language model in an environment using online RL.

## Introduction

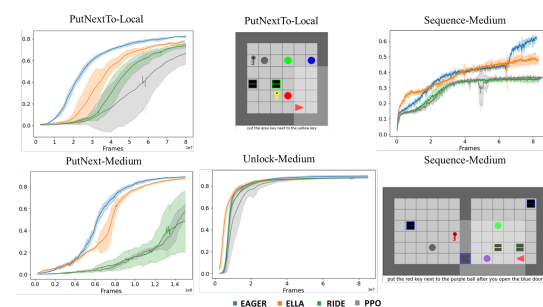
Reinforcement learning and language modelling, particularly through large language models (LLMs), are very active areas of research which, until recently, were rarely brought together. In the course of our research we have been interested in the mutual benefits of interaction between language and RL. The former provides the latter with useful structures for understanding the environment. The latter enables the former to become better embedded in the environment.

## Using Language to shape RL exploration

Reinforcement learning (RL) in long horizon and sparse reward tasks is notoriously difficult and requires a lot of training steps. A standard solution to speed up the process is to leverage additional reward signals, shaping it to better guide the learning process. In the context of language-conditioned RL, the abstraction and generalisation properties of the language input provide opportunities for more efficient ways of shaping the reward. We leverage this idea and propose an automated reward shaping method EAGER where the agent extracts auxiliary objectives from the general language goal. These auxiliary objectives use a question generation (QG) and question answering (QA) system : they consist of questions leading the agent to try to reconstruct partial information about the global goal using its own trajectory. When it succeeds, it receives an intrinsic reward proportional to its confidence in its answer (as illustrated in figure 3.1). This incentivizes the agent to generate trajectories which unambiguously explain various aspects of the general language goal. Our experimental study shows that this approach, which does not require engineer intervention to design the auxiliary objectives, improves sample efficiency by effectively directing exploration 3.2.



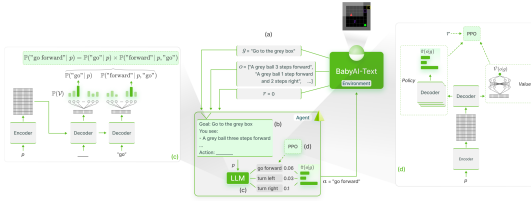
**FIGURE 3.1** – During training, the agent uses the goal to generate relevant questions using its question-generation module QG. Then, it attempts at answering them from current trajectories at each step with its question-answering module QA, by looking at the trajectory. When it succeeds, it obtains an intrinsic reward proportional to its confidence in its answer. Then it removes the answered questions from the list of questions. This incentivizes the agent to produce trajectories that enable to reconstruct unambiguously partial information about the general language goal, enabling to shape rewards and guide learning.



**FIGURE 3.2** – Average reward for EAGER and baselines for four tasks, with error regions to indicate standard deviation over four random seeds. For the PutNext-Local and Sequence-Medium tasks, we give an example of possible tasks with environment at time step 0.



## Using RL to ground Large Language Models



**FIGURE 3.3 – The GLAM method :** we use an LLM as agent policy in an interactive textual RL environment (BabyAI-Text) where the LLM is trained to achieve language goals using online RL (PPO), enabling functional grounding. (a) BabyAI-Text provides a goal description for the current episode as well as a description of the agent observation and a scalar reward for the current step. (b) At each step, we gather the goal description and the observation in a prompt sent to our LLM. (c) For each possible action, we use the encoder to generate a representation of the prompt and compute the conditional probability of tokens composing the action given the prompt. Once the probability of each action is estimated, we compute a softmax function over these probabilities and sample an action according to this distribution. That is, the LLM is our agent policy. (d) We use the reward returned by the environment to finetune the LLM using PPO. For this, we estimate the value of the current observation by adding a value head on top of our LLM. Finally, we backpropagate the gradient through the LLM (and its value head)

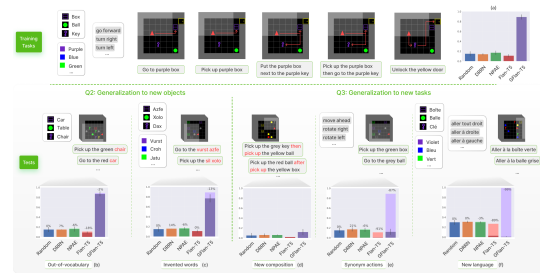
The recent rise of Transformer-based Large Language Models (LLMs) trained on massive text datasets has led to models exhibiting impressive capabilities (e.g. natural language generation, question answering, reasoning, translation...). Recently, LLMs were shown to also capture aspects of the physical rules in our world, e.g. about space, colors or even affordances between bodies and objects.

However, LLMs are known to suffer from a lack of grounding (i.e. connecting their inner representation of language to a world) which prevents them from properly dealing with the meaning of concepts and their direct use to solve tasks in interactive environments. Indeed, alignment between statistical structures in such LLMs and environments can be very limited, or even sometimes entirely wrong. This is partly due to 1) a training process (predicting next words) that is not directly incentivized to solve problems in an environment, 2) lack of abilities to intervene in the environment to identify causal structures; 3) lack in abilities to learn based on data collected as a result of interacting with the environment.

Focusing on such a functional competence, we propose to use an LLM as the policy interacting with a textual environment (i.e. a textual description of the scene is provided by the environment and possible actions are text commands) for decision-making problems. Using Reinforcement Learning (RL) to finetune the LLM to make it solve various tasks in this environment, our method named GLAM “functionally grounds” the LLM on the environment 3.3. That is, grounding the dynamics and physical rules of an environment to solve problems

and obtain, in the end, an operational LLM able to use natural language to solve tasks in this interactive environment.

Applying GLAM to Flan-T5 780M shows how such an LLM can be functionally grounded and be able to solve the tasks in an interactive textual environment. The resulting LLM also exhibits strong generalization when exposed to variations in the objects it must interact with 3.4. Finally, we show that incremental intervention using RL is key by comparing it to passive Imitation Learning. Our results highlight that our GLAM method leads to both better results and generalization even when Imitation Learning is performed using an optimal policy.



**FIGURE 3.4 – Generalization tests :** We train all agents on a mix of 5 different tasks and evaluate their generalization abilities on 1000 test episodes (also containing a mix of these 5 tasks) (a). We compare them to two baselines : an agent choosing actions randomly (Random) and the zero-shot Flan-T5 (without any finetuning). We then perform several generalization studies by (b) substituting object names out-of-vocabulary names, (c) substituting objects and colors by invented words, (d) testing a new composition of tasks, (e) substituting actions by synonyms and (f) translating the whole environment to French for the Go To task. For each agent, we plot its mean success rate over 2 seeds along with the confidence interval and the delta with performance on the same task without any change (except for (d), on which no baseline result can be provided as this task is completely new).



# EXPLORING THE IMPACT OF NEURAL NETWORK ARCHITECTURE ON DECISION-MAKING PROCESSES

Naomi CHAIX-EICHEL

Direction : Nicolas P. Rougier  
Co-Direction : Thomas Boraud

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

In this thesis, we investigated the impact of neural network architecture design on decision-making processes in various scenarios, including action selection, navigation, and risk assessment. To gain deeper insights into this question, we built computational models with an approach that did not impose rigid assumptions about the architecture of decision-making networks. Instead, we exploit models based on the reservoir computing paradigm that allows to solve complex task based on a random architecture. This approach enabled us to investigate the impact of neural network architecture on decision-making processes without bias towards preconceived notions of network structure.

## Exploring the impact of network structure in action selection processes

We delved into the study of decision-making, specifically in the context of action selection. We first followed the classical method of building models by drawing direct comparisons with the biological brain, notably the Basal Ganglia (BG), in terms of structural configuration. However, we later chose to pursue a different approach by employing reservoir computing to build our models. This alternative method offers us the flexibility to explore various architectures with fewer constraints.

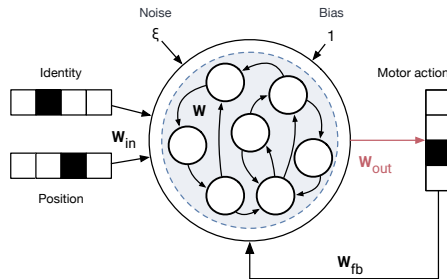


FIGURE 4.1 – Model architecture with a motor output (direction of movement). The black arrows are fixed and the red is plastic.

We first explore the structure and evolutionary preservation of the basal ganglia, followed by an introduction to the Meropi [1] and Sequential Meropi models that draw inspiration from it. The limitations and constraints of these models are critically examined. We then introduce Echo State Networks (ESN) as in 4.1, their random architecture, and their ability to solve complex tasks. We study the transition from random to ordered ESN architectures and examines a study on the role of structure in time-constrained decision tasks [2] as in 4.2 . The potential of using genetic algorithms to find the optimal structure is discussed, although this part is not

yet completed. The section concludes with a discussion on the limitations and constraints of the models explored, providing a comprehensive overview of the role of reservoir computing in understanding the structure of the basal ganglia.

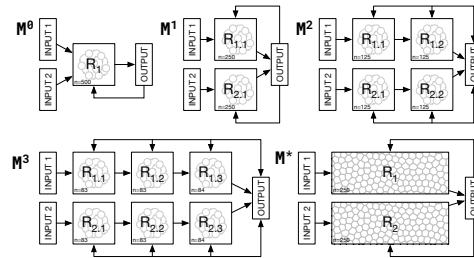
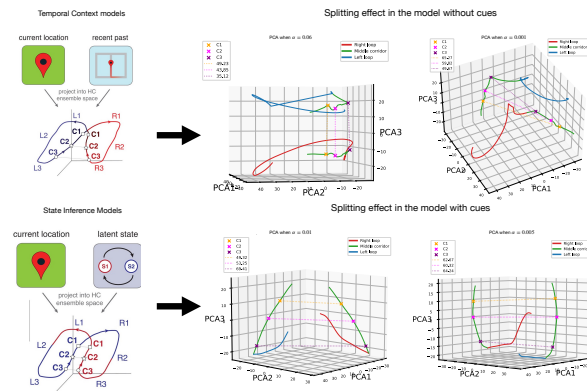


FIGURE 4.2 – Model are composed of one to several chained ESN, all connected to the readout, all receiving feedback from the output.  $M^0$  : Regular ESN.  $M^1$  : Dual pathway each made of a single ESN.  $M^2$  : Dual pathway each made of two chained ESNs.  $M^3$  : Dual pathway each made of three chained ESNs.  $M^*$  : Dual pathway each made of one continuous ESN

## Exploring the impact of network structure in navigation tasks.

In this section, we aim to explore the impact of network structure on the emergence of critical hippocampal cells, such as place cells and grid cells and splitter cells [3], which play essential roles in spatial navigation. To achieve this objective, we focus on developing a model based on reservoir computing and capable of performing a continuous navigation task, allowing us to identify the presence of these cells within the model and draw conclusions about the influence of neural architecture on their emergence. The review of Duvelle et al [3] allowed us to have a comprehensive understanding of existing hippocampal models and the emergence of hippocampal cells within them. We then introduce a study on splitter cells inside a random but embodied

computational model offering insights into the role of network structure in the emergence of these crucial cells. By connecting the findings of this study with the broader context of hippocampal models, we aim to provide a deeper understanding of the relationship between neural architecture and the development of key hippocampal cells. We also discuss the limitations of the outside-in method, which is commonly used in the field.



**FIGURE 4.3 – 3D PCA analysis on the internal states of the reservoir allowing to put in evidence the presence of hippocampal cells.**

### ***Investigating the Influence of Network Architecture on Risk Attitude in Decision-Making Under Uncertainty***

The aim of this section is to explore how network architecture impacts risk attitude in decision-making under uncertainty, using the prospect theory framework (PT) as a foundation. We begin by introducing the PT, which is a behavioral economic theory that describes how individuals evaluate potential losses and gains in the context of risk and uncertainty. We also examine the attitude toward risks in macaque decision-making, as studying non-human primates can provide valuable insights into the evolution and neural basis of risk-related behaviors. We then proceed to discuss a Prospect-Theory-aligned computational model, which is designed to simulate and better understand the underlying mechanisms that contribute to risk attitude in decision-making. By analyzing the model's structure and performance, we aim to uncover the role of network architecture in shaping risk-related behaviors and preferences.

### ***Conclusion***

In conclusion, we reflect on the unique approach we have taken in comparison to the classical method of drawing direct inspiration from brain architecture. Our findings demonstrate that simpler architectures may be sufficient to capture essential features of neural systems, challenging the prevailing assumption that complex biological structures are always necessary to accurately replicate neural phenomena.

This discovery raises questions about the methodology employed in computational neuroscience research, particularly concerning the assumptions that underlie the

modeling of neural systems. By highlighting the potential of simpler architectures, we encourage a reevaluation of the current practices in the field and emphasize the importance of exploring alternative approaches to understanding neural phenomena.

### **Bibliographie**

- [1] Topalidou, M., Kase, D., Boraud, T., & Rougier, N. P. (2018). A computational model of dual competition between the basal ganglia and the cortex. *eneuro*, 5(6)
- [2] Chaix-Eichel, N., Venugopal, G., Boraud, T., & Rougier, N. P. (2024). Exploring the role of structure in a time constrained decision task. *arXiv preprint arXiv :2401.10849*.
- [3] Duvelle, É., Grieses, R. M., & Van der Meer, M. A. (2023). Temporal context and latent state inference in the hippocampal splitter signal. *ELife*, 12, e82357.
- [4] Chaix-Eichel, N., Dagar, S., Lanneau, Q., Sobriél, K., Boraud, T., Alexandre, F., & Rougier, N. P. (2022). From implicit learning to explicit representations. *arXiv preprint arXiv :2204.02484*.

# IDENTIFIER ROBUSTEMENT DES BIOMARQUEURS DE CANCERS ET EXPLORER LEUR INTÉRÊT CLINIQUE

Elsa CLAUDE

Direction : Patricia THÉBAULT

Co-Direction : Arnaud DROIT

Encadrement : Raluca URICARU et Mickaël LECLERCQ

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Département Médecine Moléculaire, Centre de Recherche CHU de Québec, Université Laval, Québec, QC, Canada

**L**es données omiques sont un grand atout pour la recherche biomédicale et notamment pour identifier des biomarqueurs de maladies complexes comme le cancer. Les dernières décennies ont vu l'émergence de multiples méthodes dont une grande partie se basent sur du machine learning pour explorer ces données et trouver des pistes d'investigation à privilégier. Malheureusement, suivant la méthode appliquée et malgré un même jeu de données, les conclusions à tirer peuvent grandement varier. Dans un premier temps nous avons proposé de tirer partie de l'hétérogénéité des algorithmes de machine learning pour proposer une méthodologie hybride ensembliste de sélection de variable appliquée à la recherche de biomarqueurs du cancer à partir de données de transcriptomique. Dans un second temps pour favoriser l'exploration et la communication de ces résultats à la communauté de recherche biomédicale nous développons un service web implémentant notamment une visualisation par graphe dynamique enrichi d'informations d'annotations biologiques.

## Objectif

Le cancer continue de causer de nombreuses victimes chaque année malgré les progrès de la recherche biomédicale. Il est donc crucial de poursuivre les efforts pour mieux comprendre, diagnostiquer, surveiller et traiter cette maladie. Les données omiques fournissent des informations moléculaires essentielles, mais leur analyse présente des défis techniques tels que le "fléau de la dimension", où le nombre de variables dépasse largement le nombre d'observations [1]. Des méthodes de sélection de variables adaptées sont nécessaires pour identifier les variables pertinentes parmi celles disponibles. Les méthodes de sélection de variables basées sur l'apprentissage automatique ont gagné en popularité, mais manquent encore de robustesse [2]. Ainsi, les approches ensemblistes sont devenues courantes car elles permettent de prendre en compte des perturbations des données (ex : ré-échantillonnage des données) ou fonctionnelles (ex : emploi de multiple algorithmes de sélection de variable) [3]. Pour aller plus loin, les approches hybrides ensemblistes de sélection de variable (HEFS) qui sont de plus en plus utilisées mettent en place une association de ces deux types de perturbation [4]. Leur emploi est néanmoins encore récent et les diverses étapes qui composent sa mise en place sont peu critiquées. Ainsi, dans cette thèse notre premier objectif vise à répondre au besoin de méthodologie pour la mise en place d'approches HEFS robustes pour l'identification de biomarqueurs de maladies complexes comme le cancer à partir de données de transcriptomique. Le second objectif consiste à évaluer

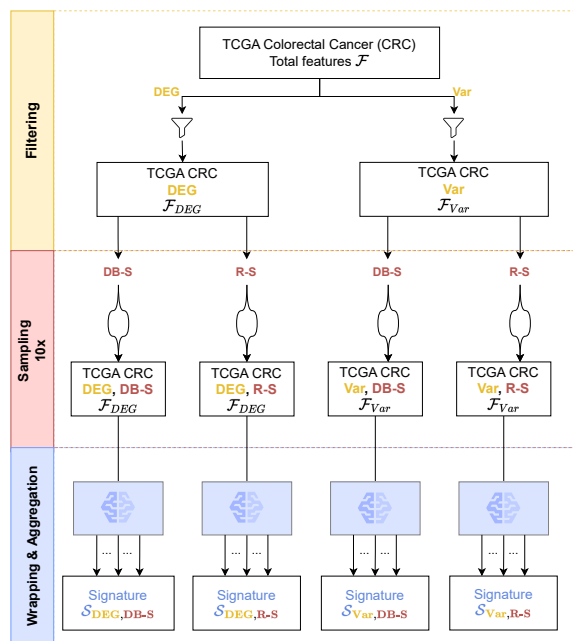
cliniquement les biomarqueurs identifiés, notamment dans un contexte pan-cancer, afin de concentrer les efforts de recherche biomédicale sur des pistes prometteuses. Un service web interactif sera développé pour visualiser dynamiquement les phénotypes cancéreux, les biomarqueurs et les voies biologiques potentielles.

## Contributions

### Objectif 1

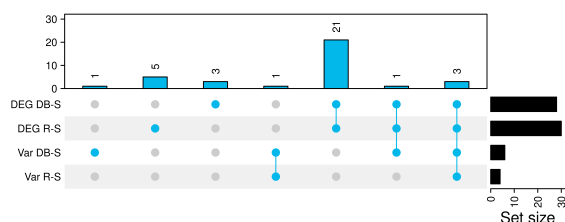
**Matériel et Méthode.** Les données d'expression génique (RNA-seq) provenant d'échantillons sains et de stade IV du cancer colorectal ont été collectés via le projet The Cancer Genome Atlas (TCGA). Nous disposons alors de 138 échantillons au total pour approximativement 40 000 variables. Notre approche HEFS (Figure 5.1) comprend une étape de diminution des dimensions par filtre de variance (Var) ou par analyse de gènes différentiellement exprimés (DEG). À la suite de cela, 10 paires de jeux de données d'entraînement et de test sont produits par une méthode d'échantillonnage par holdout répété couplé à une méthode de distribution équilibrée (distribution-balanced) ou de stratification aléatoire. Ces divers ensembles de données sont soumis à une même étape d'entraînement d'algorithme d'apprentissage automatique par *wrapper*. Enfin une procédure d'aggrégation est mise en place pour permettre la mise en commun conditionnée des diverses listes de variables utilisées par les modèles pour permettre la différenciation des échantillons sains et de stade IV du cancer en cause. Nous avons envisagé 4 scénarios d'HEFS dont les résultats sont des listes robustes

de variables (et donc de biomarqueurs potentiels).



**FIGURE 5.1** – Vue d'ensemble des multiples scénarios d'approches hybrides ensemblistes de sélection de variable, des étapes de filtres et d'échantillonnage à la sélection de caractéristiques avec des modèles d'apprentissage enveloppants et une étape finale d'agrégation.

**Résultats.** Les divers scénarios d'HEFS ont permis l'obtention de courtes signatures de biomarqueurs potentiels longues de 4 à 30 variables depuis les 40 000 variables initiales (Figure 5.2). Les signatures les plus petites correspondant aux scénarios HEFS Var ont un aussi fort potentiel de séparation des échantillons sains et de stade IV que les variables obtenues par scénario DEG sur un jeu de données externe (GSE50760). L'intérêt clinique de ces diverses variables et plus précisément des trois communes aux 4 scénarios a été vérifié après une analyse bibliographique complète. Notre étude montre que les approches d'HEFS sont fiables mais doivent encore faire l'objet d'une attention particulière de l'utilisateur quant aux diverses étapes de conception afin de limiter les biais méthodologiques.



**FIGURE 5.2** – Distribution des variables dans les signatures obtenues par les 4 scénarios d'HEFS étudiés.

La complétion de ce premier objectif a fait l'objet d'une soumission au journal Nucleic Acids Research Genomics and Bioinformatics et est actuellement en procédure de révision.

## Objectif 2

L'objectif 2 est en cours de développement. Sa réalisation s'appuie sur l'analyse de données d'une quinzaine de cancers différents avec des comparaisons de classes d'échantillons variées (Sain vs Stade I, Stade I vs Stade IV etc.). Cette analyse permise par une approche d'HEFS développée pendant le premier objectif vise à mettre en relief des listes restreintes et robustes de biomarqueurs potentiels. Pour pouvoir explorer ces biomarqueurs nous permettons à l'utilisateur de parcourir un large graphe non orienté où les noeuds sont des gènes ou des voies biologiques et les arêtes peuvent définir des liens entre gènes de signatures différentes ou des liens entre des gènes et des voies biologiques. Le graphe est dynamique et les interactions avec ce dernier doivent permettre à l'utilisateur d'identifier des signatures, les comparer par des liens directs (intersection de liste de gènes) ou indirects (par voies biologiques).

## Bibliographie

- [1] Bishop,C.M., Hinton,G., Bishop,C.M. and Hinton,G. (1995) Neural Networks for Pattern Recognition Oxford University Press, Oxford, New York.
- [2] Long,N.P., Park,S., Anh,N.H., Nghi,T.D., Yoon,S.J., Park,J.H., Lim,J. and Kwon,S.W. (2019) High-Throughput Omics and Statistical Learning Integration for the Discovery and Validation of Novel Diagnostic Signatures in Colorectal Cancer. Int. J. Mol. Sci., 20, 296.
- [3] Bolón-Canedo,V. and Alonso-Betanzos,A. (2019) Ensembles for feature selection : A review and future trends. Inf. Fusion, 52, 1-12.
- [4] Colombelli,F., Kowalski,T.W. and Recamonde-Mendoza,M. (2022) A hybrid ensemble feature selection design for candidate biomarkers discovery from transcriptome profiles. Knowl.-Based Syst., 254, 109655.

# REACHABILITY IN TEMPORAL GRAPHS

Timothée CORSINI

Direction : Arnaud CASTEIGTS

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

**T**emporal graphs are a tool used to represent a graph evolving over time. In these, the notion of reachability between two vertices differs from static graphs since there may not be a path at a given time instant nor should a path in the underlying graph implies one in the temporal one. This short document aims to highlight some of my work regarding temporal graphs and reachability problems.

## Temporal graphs

We define a temporal graph as a labeled graph  $\mathcal{G} = (V, E, \lambda)$  where  $V$  is a finite set of vertices,  $E \subseteq V \times V$  a set of undirected edges, and  $\lambda : E \rightarrow 2^{\mathbb{N}}$  a function assigning at least one time label to every edge, interpreted as presence times. The graph may also be seen as a sequence of static graphs called snapshots (see Figure 6.1 for an example). We call the graph  $(V, E)$  without the labels the footprint of  $\mathcal{G}$ .

A fundamental aspect of temporal graphs is reachability. The reachability of a temporal graph  $\mathcal{G}$  is characterized by the existence of temporal paths or journeys, a path which traverses edges in chronological order. When every vertex may reach each other, the graph is said to be temporally connected.

From a temporal graph  $\mathcal{G}$ , we can build a static directed graph called the reachability graph of  $\mathcal{G}$  where an arc  $(u, v)$  exists if and only if  $u$  can reach  $v$  in  $\mathcal{G}$ . Observe that  $\mathcal{G}$  is temporally connected if and only if its reachability graph is a clique.

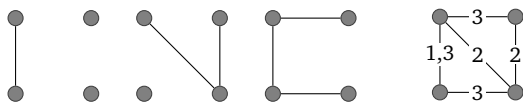


FIGURE 6.1 – The snapshots of the temporal graph and the labeled representation.

## Temporal graphs settings

When studying journeys, many consider that either a path crosses at most one edge at a time or any number of edges, this distinction is what we call strict journeys for the former and non-strict journeys for the latter. This leads to a lot of independent results on some problems as the behaviour of the graph may differ greatly whether we consider strict or non-strict journeys.

A way to reconcile the two notion is by restricting the structure of the temporal graph to a proper labeling, two adjacent edges cannot share any common label, in other words, no edges may appear next to each other, leading to every possible journey being strict. A temporal graph with such labeling is called proper, finding NP-hard results on such graphs can be expanded to non-proper graphs regardless of strictness.

Finally, another property to consider is whether an edge may be present at multiple time steps, if not, every edge has a single label. Such graphs, called simple temporal graphs are much more restricted than general temporal graphs while still encapsulating their complexity.

In many papers, these three properties, strictness, properness and simpleness are used but not always compared to each other in terms of results, which was the motivation of my work with Arnaud in [4] which represents these properties and their implications, an overview of the settings induced by the properties can be seen in Figure 6.2. In our work, we consider simple and proper graphs, the most restricted graphs considering the above settings while still being very complex in their own way. We call them happy graphs as any hardness results on them may be transferred to non-proper as well as non-simple graphs, but on the other hand, some positive results on a happy graph may be transferred to a general one through transformations or by finding such happy graph as a subgraph.

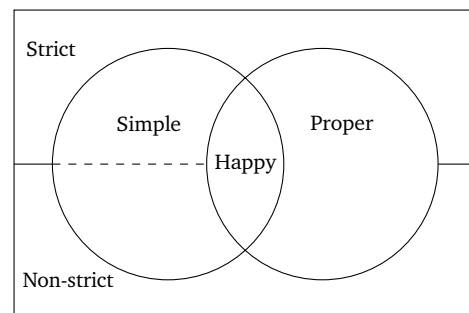


FIGURE 6.2 – Settings resulting from the three properties

## Some problems on temporal graphs

### Minimizing temporal graphs

In temporal graphs, many problems from static graphs become much more complex. For instance the problem of finding a minimum temporally connected subgraph (called the spanner problem here) is a translation from finding spanning trees, however an optimal spanner may be much bigger, it has been proved that regardless of the setting there may exists minimum spanners of size  $\Theta(|V|^2)$  [2], something much worse than the



linear size for spanning trees. Moreover, in most of the settings, it is APX-hard to find the size of the minimum spanner [1, 2]. The only exception is the happy setting where the complexity of the problem is still open.

In my thesis, we worked on a special case of temporal spanners, temporal spanning trees [3]. Interestingly enough, even if it is NP-hard to find the optimum spanner in most cases, deciding the existence of temporal tree was still open in term of complexity. However, we proved that it was still NP-hard, and so, even in the proper setting, and that it is trivially polynomial in the simple settings.

The study of the spanning tree allowed us to consider more restricted temporal paths, since a temporal tree requires each pair of vertices to use the same underlying path to reach each other, we call these paths bidirectional journeys. The existence of bidirectional journeys is implied by a spanning tree but the converse is not true and finding such bidirectional journeys can be done in polynomial time. Sadly, like temporal spanners, we showed that optimizing the size of bidirectional spanners (where every vertices must share a bidirectional journey) is NP-hard in the simple and non-strict setting, though in the case of the simple and strict setting, the problem is equivalent to limiting the graph to a clique.

A more general problem would be, considering a temporal graph, to highlight which labels or edges are mandatory to preserve the reachability graph, and minimizing the size of a temporal graph while preserving its reachability graph, this is of course at least as hard as the spanner problem since it contains it.

### Temporal graphs from reachability graphs

Instead of taking a temporal graph and trying to reduce its size, other questions considered are related to whether a static graph may lead to a temporal one with some restrictions. For instance, given a static connected graph, is it possible to label it and make it temporally connected? While this is true in the non-simple or non-strict settings, it is much harder to decide in the happy or simple and strict settings.

During my thesis we considered the following problem : given a static directed graph, can it be the reachability graph of some undirected temporal graph? Or put differently, what properties does the class of reachability graphs has?

This question has different answers depending on the setting, when looking at Figure 6.3, the first graph which is a path will never be the reachability graph of some non-strict temporal one since any edge labeling of a graph with 3 vertices would lead to a journey from the first vertex to the last or the opposite case. This graph however can be reachability of some strict graph, the most simple example is the same graph labeled with 1 on each edge.

The second graph from the figure cannot exists in the simple and strict case as any temporal graph which would have such reachability would have a path as a footprint, an arc in the reachability graph from  $a$  to  $c$

would imply that the tome label on  $ab$  is lower than  $ac$ , the same argument goes for  $b$  to  $d$ , this would force an arc from  $a$  to  $d$  in the reachability graph.

Finally the last graph shows that regardless of the settings, no undirected temporal graph may have this reachability graph. No arc between  $a$  and  $c$  means that the setting must be strict and  $ab, bc$  share a single label that is the same, the same goes between  $ac$  and  $bd$ , however the arc  $(c, d)$  would mean that  $bc$  has a lower time label than  $bd$ , leading to a contradiction.

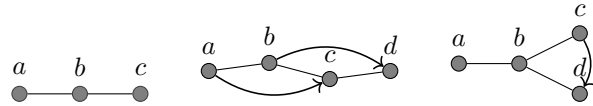


FIGURE 6.3 – Forbidden reachability from left to right : for the non-strict settings, for the simple and strict setting, for all settings

This reachability property allows us to distinguish the settings in term of which reachability graph are achievable, however the complete description of the reachability graphs achievable is still an open question [4].

### Bibliographie

- [1] Akrida, E. C., Gašieniec, L., Mertzios, G. B., and Spirakis, P. G. (2017). The complexity of optimal design of temporally connected graphs. *Theory of Computing Systems*, 61, 907-944.
- [2] Axiotis, Kyriakos, and Dimitris Fotakis. "On the size and the approximability of minimum temporally connected subgraphs." *arXiv preprint arXiv :1602.06411* (2016).
- [3] Casteigts, A., and Corsini, T. (2023). In search of the lost tree : Hardness and relaxation of spanning trees in temporal graphs. *arXiv preprint arXiv :2312.06260*.
- [4] Casteigts, A., Corsini, T., and Sarkar, W. (2024). Simple, strict, proper, happy : A study of reachability in temporal graphs. *Theoretical Computer Science*, 114434.

# DÉVELOPPEMENT D'INFRASTRUCTURE NUMÉRIQUE POUR LA SÛRETÉ FERROVIAIRE

Mayssa DARDOUR

Direction : Mohamed MOSBAH

Co-Direction : Toufik AHMED

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

**C**e projet de recherche explore l'amélioration de la sûreté et de l'efficacité dans le réseau ferroviaire à travers l'adoption des technologies de l'information et de la communication, notamment l'Internet des véhicules (IoV). Il vise à développer des solutions innovantes pour la gestion des risques aux passages à niveau et l'optimisation des opérations ferroviaires, grâce à l'utilisation de capteurs intelligents, de communications sans fil haut débit, et de l'intelligence artificielle. L'objectif est d'accroître la sécurité, réduire les retards, et améliorer l'expérience des usagers tout en conservant une approche éco-énergétique. Ce travail promet d'offrir des avancées significatives dans la gestion des infrastructures ferroviaires et la prestation de services de transport plus sûrs et plus efficaces.

Les systèmes de transports intelligents reposent sur les nouvelles technologies pour proposer de nouveaux services au réseau ferroviaire, tels que la sécurité intégrée, la gestion des actifs et la maintenance prédictive afin d'améliorer la prise de décision en temps utile pour des questions telles que la sécurité des voies aux trains ou lors des passages à niveau, les horaires et la capacité du système. L'objectif est de répondre à une demande toujours croissante de services efficaces et fiables, tout en tenant compte du facteur énergétique en exploitant des capteurs intelligents, des technologies sans fil offrant une connectivité haut débit, des nouvelles technologies de communication et des solutions logicielles intégrées. Avec l'avènement du concept de l'Internet des véhicules, une panoplie de nouvelles technologies sans fil peut être exploitée pour offrir des services temps réel liés à la sûreté. Un exemple typique d'application est d'assurer la sûreté des passages à niveau et éviter les collisions possibles avec les voitures. En effet, des aléas peuvent modifier les horaires de passage des trains, tels qu'un temps de stationnement plus long, une arrivée retardée, un incident, etc. ; pouvoir prédire les temps de stationnement ou les retards reste un problème d'actualité pour le réseau ferroviaire. Pour cela, la sécurité ferroviaire reste un souci permanent, puisque les incidents ferroviaires impliquent souvent un nombre important de personnes. Il est essentiel de maintenir et d'améliorer la sécurité, tout en réduisant les coûts.

Dans ce contexte, profiter des apports de l'IoV et les opportunités qu'il apporte est une solution attrayante. Ainsi, il est possible de proposer des applications qui vont intervenir le réseau ferroviaire et son environnement (réseau véhiculaire, etc.) et qui sont liées à la sécurité, la sûreté, l'efficacité du trafic, l'info-divertissement, etc. Un train connecté pourrait envoyer et communiquer avec l'infrastructure, les véhicules connectés, les piétons, les cyclistes,... pour envoyer des alertes ou des informa-

tions urgentes. Outre ses services de sûreté et sécurité, un train connecté offre d'autres services de divertissement à ses passagers. Pour satisfaire les exigences de toutes ces applications ayant des contraintes différentes, il faut s'appuyer sur des réseaux de communication à haut débit ininterrompus et tolérants aux pannes pour toutes les conditions géographiques, structurelles et météorologiques possibles. L'orientation actuelle est d'utiliser les nouvelles technologies de communication (Wi-Fi, le WIMAX, le LTE-R, GSM-R, ITS-G5, 5G, etc.). Nous proposons dans le cadre de ce projet d'explorer les communications et le traitement Edge pour améliorer la sécurité des croisements. Le défi est à la fois de garantir la sûreté dans un environnement critique et de réduire les calculs. Il est de plus en plus admis que les infrastructures numériques vont fournir des services de stockage proche des utilisateurs (EDGE). Ceci va permettre de réduire les accès au cloud distant et les calculs supplémentaires, et donc réduire la consommation énergétique. Cependant, il faut s'assurer de la cohérence et de la consistance des données traitées. En explorant les progrès technologiques de ces futurs réseaux, mais aussi les techniques d'intelligence artificielle pour le traitement de données de proximité, nous pensons réduire considérablement les calculs.

Plus précisément, nous proposons les étapes suivantes :

- Étudier les performances des communications ITS pour les trains (slicing, hybridité, etc.).
- Étudier la mise en oeuvre d'une architecture de communication globale entre tous les modes de transport et toutes les sources de données.
- Étudier le traitement de données et les solutions EDGE / FOG.
- Développer des algorithmes d'intelligence artificielle pour la sûreté.
- Valider les approches retenues par des techniques de sûreté de fonctionnement.

- Application : croisements et passages à niveau dans un contexte C-ITS.
- Simulation et prototypage des solutions proposées.
- Expérimentation sur le site de Ferrocampus.

#### Livrables :

- **T0+6** : Analyse critique de l'état de l'art de la sûreté des croisements et notamment des passages à niveau dans un environnement ITS.
- **T0+12** : Proposition et validation d'une architecture en intégrant les communications ITS, étude de la virtualisation.
- **T0+24** : Implémentation, études des performances, comparaison des coûts de calcul vs correction des résultats, validation.
- **T0+36** : Expérimentation dans le cadre du site Ferrocampus.

## Bibliographie

- [1] Zhou, Z., Liao, H., Wang, X., Mumtaz, S., and Rodriguez, J., "When Vehicular Fog Computing Meets Autonomous Driving : Computational Resource Management and Task Offloading," in *IEEE Network*, vol. 34, no. 6, pp. 70-76, November/December 2020, doi : 10.1109/M-NET.001.1900527.
- [2] Fraga-Lamas, P., Fernández-Caramés, T. M., & Castedo, L. (2017). "Towards the Internet of smart trains : A review on industrial IoT-connected railways." *Sensors*, 17(6), 1457.
- [3] Ma, X., et al., "Prototypes of using directional antenna for railroad crossing safety applications," 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2017, pp. 594-596, doi : 10.1109/CCNC.2017.7983182.
- [4] Uhlemann, E. (2015). "Introducing connected vehicles [connected vehicles]." *IEEE Vehicular Technology Magazine*, 10(1), 23-31.
- [5] Banerjee, S., Hempel, M., & Sharif, H. (2016). "A survey of wireless communication technologies & their performance for high speed railways."
- [6] 3GPP TR 21.914 V0.8.0 "3rd Generation Partnership Project ; Technical Specification Group Services and System Aspects ; Release 14 Description ; Summary of Rel-14 Work Items (Release 14)," September 2017.
- [7] Abdallah Moubayed, Abdallah Shami, Parisa Heidari, Adel Larabi, Richard Brunner "Edge-Enabled V2X Service Placement for Intelligent Transportation Systems," January 2020
- [8] ETSI TS 122 185 V14.3.0 - 3GPP TS 22.185 version 14.3.0 Release 14 "Service requirements for V2X services," March 2017
- [9] 3GPP TR 21.915 V0.5.0 "3rd Generation Partnership Project ; Technical Specification Group Services and System Aspects ; Release 15 Description ; Summary of Rel-15 Work Items (Release 15)," December 2018.
- [10] 3GPP TR 21.916 V0.5.0 "3rd Generation Partnership Project ; Technical Specification Group Services and System Aspects ; Release 16 Description ; Summary of Rel-16 Work Items (Release 16)," July 2020.
- [11] D. Kehagias, C. Sougles, M. Gkemou, (CERTH), A. Steccanella (CRF), "SAFE and green Sensor Technologies for self-explaining and forgiving Road Interactive applications, WP2 : System architecture and specifications, Grant Agreement Number : 723211," August 2017.
- [12] Younes Bouchaala, "Gestion des messages de sécurité dans les réseaux VANET," Thèse de doctorat de l'Université Paris-Saclay préparée à l'UVSQ, August 2018
- [13] Mukesh Kumar Saini, Abdulmotaleb El Saddik, Abdulhameed Alelaiwi, *ACM Computing Surveys* 48(2) :1-40, DOI :10.1145/2817552 "How close are we to realizing a pragmatic VANET solution ? A meta-survey," November 2015.



# STARONNX : A DYNAMIC SCHEDULER FOR LOW LATENCY AND HIGH THROUGHPUT INFERENCE ON HETEROGENEOUS RESOURCES

Jean-François DAVID

Direction : Olivier BEAUMONT

Encadrement : Lionel EYRAUD-DUBOIS et Samuel THIBAUT

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

Our study delves into the StarONNX project, which aims to create a dynamic scheduler to enhance latency and maximize throughput for deep neural network (DNN) inference across varied computing resources. We strive to efficiently execute DNN models on a broad spectrum of processors, including CPUs and hardware accelerators, focusing on optimizing latency and throughput. We introduce StarONNX, which merges ONNX Runtime with StarPU for optimal task distribution and resource management in heterogeneous architectures. This method uses ONNX Runtime for effective deep learning model execution and StarPU for resource orchestration with advanced scheduling techniques to improve computation and communication overlap. A key achievement is our ability to split a CNN model for concurrent operation into GPUs and CPUs, utilizing all resources to boost throughput without sacrificing latency. Our findings indicate that integrating ONNX Runtime with StarPU introduces negligible overhead and outperforms the leading inference server, Triton, regarding resource use and latency reduction. Our approach assesses performance through throughput, inference task batch sizing, and the system's ability to process requests with minimal latency. We also investigate model partitioning strategies to optimally employ CPUs and GPUs, aiming to minimize the maximum processing, transfer, and CPU times for segments of a partitioned model.

## Introduction

Deep Neural Network (DNN) inference, crucial for AI applications, faces challenges on platforms ranging from embedded devices to large servers [1]. These challenges include processing complex data streams in real-time, achieving high throughput and low latency, and dealing with memory constraints and DNN parallelism. Traditional scheduling methods like Round Robin must be revised due to the heterogeneous nature of resources (CPUs, GPUs, FPGAs) and DNNs' complex dependency graphs. This study integrates ONNX Runtime with dynamic schedulers like StarPU[2] to meet DNN inference demands, focusing on parallel processing and efficiency to improve throughput and latency. We introduce a method for efficient DNN execution on varied platforms. We demonstrate our system's benefits over conventional approaches, addressing resource diversity and task parallelism to enhance DNN inference efficiency and scalability.

## StarONNX

### Integrating ONNX into StarPU

We assume that the neural network model is split into several parts. Each part can be optimized by ONNX Runtime for all available resources, e.g. with operator fusion, before the inference starts. In StarONNX, each of these parts corresponds to a StarPU task, and StarPU

is responsible for efficiently managing the dependencies and data transfers between tasks. Figure 8.1 shows the process of an inference in StarONNX for a neural network divided into two parts.

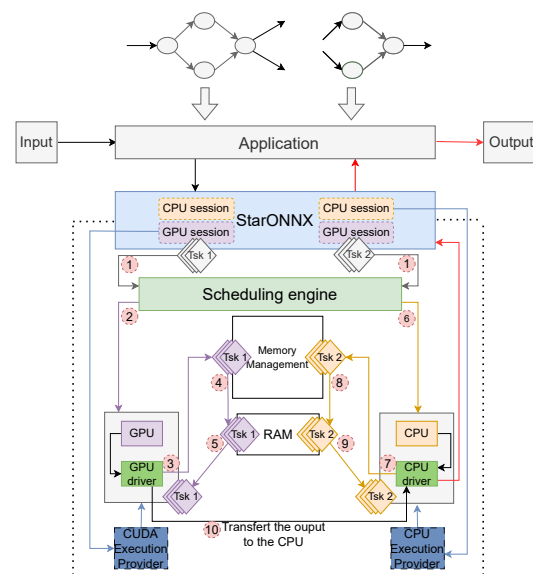


FIGURE 8.1 – An example of how StarPU works with ONNX Runtime on heterogeneous processors and a sliced model.

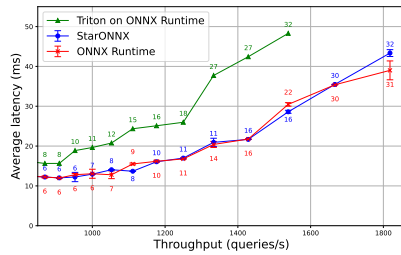


FIGURE 8.2 – StarPU overheads by integrating ONNX Runtime and comparison with Triton.

### Performance Evaluation

Let us compare the performance of Native ONNX, StarONNX, and the Triton inference server for a single GPU and without model partitioning. To compare the performance of different inference systems, we specify an arrival rate for inference tasks and search for the batch size that (i) allows inference to be performed with the imposed arrival rate (ii) while minimizing latency. The results for the CNN GoogLeNet [3] model are shown in Figure 8.2. For the Native ONNX results, we implemented by hand pinned memory, task concurrency, and prefetching for overlapping communication and computation. Execution is asynchronous, copying data from the host to the driver on another stream. Therefore, for the purposes of fair comparisons, we introduce the same advantages as StarPU. Interestingly, using StarPU through StarONNX allows users to seamlessly benefit from sophisticated communication and memory management. The results shown in Figure 8.2 illustrate that the integration of ONNX Runtime into StarPU does not introduce significant overhead, and trace analysis verifies that the utilization of computational and communication resources is indeed very close to optimal. Conversely, Triton introduces significant overhead compared to both optimized ONNX and StarPU.

### Inference with Partitioned Neural Networks

StarONNX efficiently utilizes GPUs, handling inference requests at an optimal frequency to effectively overlap computation and communication while managing memory allocations well. It optimizes resource use across heterogeneous systems through intelligent scheduling, prioritizing tasks based on their fastest processing route considering transfer times.

We demonstrate how CPUs and GPUs are effectively leveraged. A simple approach lets inferences run on either without code changes, blending CPU and GPU capabilities. This works well until the GPU's processing time matches batch creation times, directing overflow to the CPU, which increases maximum latency.

To curb latency, we split the model in two, distributing parts between the CPU and GPU for balanced load and resource use. This reduces latency compared to CPU-only processing. Effective partitioning considers computation and transfer times and a Python script identifies optimal splits using ONNX Runtime.

In Figure 8.3, we showcase a partitioned GoogLeNet model's performance, employing the HEFT algorithm for scheduling. The GPU manages alone under ideal conditions (850 to 1100 queries/s). Beyond this, CPU assistance is gradually enlisted, reaching complete GPU-CPU processing at 2000 queries/s, enhancing throughput (2000 vs. 1800 queries/s with GPU alone).

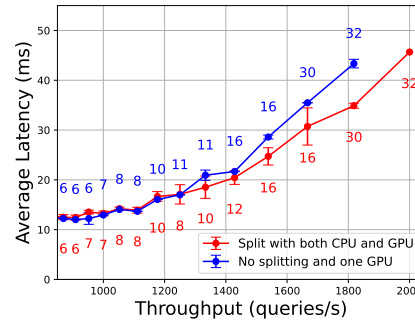


FIGURE 8.3 – Ability of StarONNX to handle heterogeneous resources for throughput.

### Conclusion

In conclusion, the synergy between ONNX Runtime and StarPU enables optimal task distribution and efficient resource management, improving the processing of complex calculations on different processors, including CPUs and hardware accelerators. StarONNX efficiently exploits GPU and CPU resources and demonstrates the effectiveness of partitioning models to balance the load and minimize latency.

This research highlights the potential of StarONNX to improve the efficiency and scalability of DNN inference, offering prospects for future advances in AI and high-performance computing.

### Bibliographie

- [1] Malik, Babur Hayat and Amir, Mehwashma and Mazhar, Bilal and Ali, Shehzad and Jalil, Rabiya and Khalid, Javaria. Comparison of task scheduling algorithms in cloud environment. Int. Journal of Advanced Computer Science and Applications
- [2] Augonnet, Cédric and Thibault, Samuel and Namyst, Raymond and Wacrenier, Pierre-André. StarPU : a unified platform for task scheduling on heterogeneous multicore architectures. 15th International Euro-Par Conference.
- [3] Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew. Going deeper with convolutions. IEEE conference on computer vision and pattern recognition.

# ANALYSE AUTOMATISÉE DE BINAIRES À LA RECHERCHE DE VULNÉRABILITÉS MATÉRIELLES

Theo DE CASTRO PINTO

Direction : Antoine Rollet

Co-Direction : Grégoire Sutre

Encadrement : Ireneusz Tobor

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

Les cartes à puce et autres appareils embarqués mettent les développeurs face à un problème de taille : les attaques physiques. Parmi celles-ci, certaines visent à modifier le comportement de la puce (temporairement ou non) grâce à une perturbation physique (laser, glitch ou autre). Ces dernières, qui peuvent être multiples au cours d'une seule exécution, sont difficiles à détecter et même si des contre-mesures existent, elles peuvent être insuffisantes ou disparaître à la compilation. Pour ces raisons une analyse des binaires générés est nécessaire, afin de détecter la présence (ou non) de vulnérabilités face à ces attaques. Cette analyse est généralement effectuée à la main. Ce papier présente un outil d'aide à la recherche de vulnérabilités matérielles pour du code binaire. Il s'appuie sur l'outil Binsec. Il permet de détecter des vulnérabilités dans un binaire en fonction d'un schéma d'attaque et d'une cible à atteindre.

## Introduction

En informatique, une vulnérabilité est une faille de sécurité exploitable par un attaquant et pouvant mener à l'altération du comportement du logiciel ou à la divulgation d'informations secrètes. Il existe des vulnérabilités dites logicielles (buffer overflow, ROP, JOP [1], ...) et des vulnérabilités dites matérielles, ce sont ces dernières qui sont considérées dans ce papier. Ces vulnérabilités peuvent être exploitées via des attaques par injection de faute (en utilisant des LASER, des glitch de tension, ...).

Une attaque matérielle peut avoir de nombreuses conséquences. Il est possible de classer ces effets en quatre catégories atomiques : les bit set (un bit du programme est mis à 1), les bit reset (un bit du programme est mis à 0), les bit flip (un bit du programme est inversé) et le rejeu d'une instruction précédente. Ces différents effets peuvent être combinés ce qui donne lieu à des attaques plus sophistiquées, par exemple, un branchement conditionnel inversé, une instruction transformée ou sautée. La conséquence de tout cela est que des comportements non-attendus peuvent être ajoutés via des attaques par injection de faute.

Il est nécessaire de limiter les vulnérabilités matérielles dans les programmes, notamment embarqués, cependant c'est une tâche ardue. Les développeurs doivent imaginer un maximum d'attaques possibles afin d'intégrer des contre-mesures<sup>1</sup> qui pourront limiter le nombre de vulnérabilités exploitables. Actuellement, les évaluations consistent en des revues de code (source en conjonction avec le code assembleur fourni par le client) à la main (faites par des experts) et ces dernières ont le désavantage d'être très longues et difficiles. Il

y a donc un besoin d'analyse automatisée dans ce milieu. L'objectif de ce projet est de développer un outil permettant d'assister les experts dans leur recherche de vulnérabilités matérielles et d'en automatiser une grande partie.

## Exécution symbolique

Notre méthodologie est une extension de celle proposée par Lazart [2]. Ce dernier est un outil qui analyse du code LLVM (compilé à partir du code source) afin d'y détecter des vulnérabilités par injection de faute. Nous avons donc développé un outil qui applique et étend cette méthodologie aux binaires, notamment via l'utilisation de Binsec [3], un outil d'analyse de binaire via l'exécution symbolique. Le but est de tester si un mauvais état du programme peut être atteint en autorisant un certain nombre d'attaque matérielle.

Cependant, l'exécution symbolique est limitée par le passage à l'échelle. Cela vient du fait que l'exécution symbolique est une stratégie visant à explorer tous les chemins d'exécution possibles. Cette exploration est guidée par une stratégie, classiquement : DFS (Depth-First Search), BFS (Breadth-First Search) et NURS (Non-Uniform Random Search). Ces stratégies n'explorent pas le programme en fonction de l'état du programme à atteindre. Nous avons donc mis au point une nouvelle stratégie d'exploration basée sur l'algorithme A\* [4] et l'avons testée via Binsec sur différents programmes. Les résultats sont présentés en Figure 9.1, chaque paquets de colonnes correspond à un programme [5] et chaque colonne de gauche à droite correspond à une stratégie (dans l'ordre : DFS, BFS, NURS, A\* et notre heuristique qui est une amélioration de A\*). Clairement,

1. Des sécurités permettant de détecter des attaques en faute.

notre nouvelle stratégie permet d'atteindre les états du programme voulus bien plus rapidement que les autres stratégies classiques.

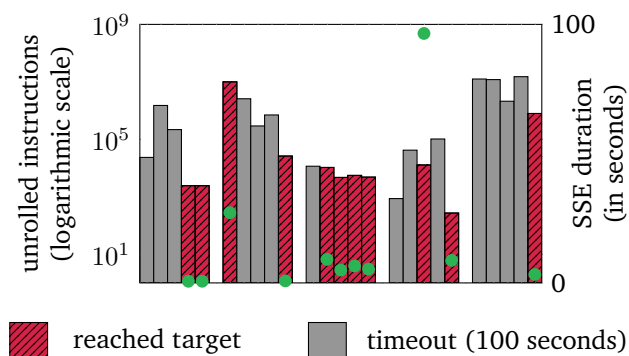


FIGURE 9.1 – Experimental results obtained with Binsec : bars represent the number of unrolled instructions in a logarithmic scale and dots represent the SSE duration in seconds.

## Injection de faute

Afin de simuler des comportements indésirables lors de l'exécution symbolique nous proposons la méthodologie suivante :

1. Transformer le binaire en un langage intermédiaire (DBA) qui correspond à un automate.
2. Ajouter des mutations dans cet automate, ce qui correspond à ajouter des transitions fautées dans l'automate.
3. Explorer symboliquement l'automate (grâce à Binsec) afin d'atteindre une cible, c'est-à-dire une ligne du programme avec comme pré-condition qu'elle n'était pas atteignable sans faire un certain nombre de faute.

Des exemples de cette méthodologie peuvent être vu en Figure 9.3, l'automate correspond au code en Figure 9.2. Les transitions en pointillées rouges sont des transitions fautées, ici le modèle de faute choisi est de pouvoir effectuer des branchements même lorsque les conditions ne sont pas remplies.

```
1 cmp r0, r1
2 bne 4
3 call critical_function
4 halt
```

FIGURE 9.2 – Code assembleur non sécurisé

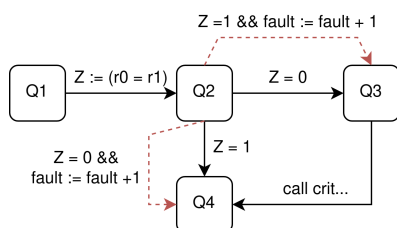


FIGURE 9.3 – Automate avec mutations correspondant au code en listing 9.2

La difficulté majeure de cette approche est de spécifier quels états sont censés être atteignables sans injection de faute. Pour ce faire, la solution envisagée est de demander à un expert travaillant sur le programme à analyser de spécifier des comportements indésirables (e.g : telle ligne de code n'est pas atteignable si tel registre vaut 0). Cette solution n'est pas optimale car la spécification d'un tel comportement n'est pas triviale, et un expert du programme n'a pas nécessairement de compétences en formalisation. Pour cette raison, il faut aussi développer un langage de haut niveau afin de permettre à un expert de facilement spécifier les comportements indésirables que Binsec doit détecter. Avec cette stratégie, il est possible d'obtenir des résultats sensiblement meilleurs que ceux avancés par Lazart : sur un exemple classique de l'état de l'art, Wookey [6], Lazart détecte deux vulnérabilités, et l'outil que nous avons développé en détecte trois, dont une non documentée par Lazart. Aussi, lors du développement de notre outil, Lazart utilisait uniquement un modèle de faute se basant sur l'inversion de branchements. L'outil développé quant à lui peut travailler avec n'importe quel modèle de faute (mais pas nécessairement de manière efficace).

## Bibliographie

- [1] Lee, Yongje, et al. "Integration of ROP/JOP monitoring IPs in an ARM-based SoC." 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, 2016.
- [2] Potet, Marie-Laure, et al. "Lazart : A symbolic approach for evaluation the robustness of secured codes against control flow injections." 2014 IEEE Seventh International Conference on Software Testing, Verification and Validation. IEEE, 2014.
- [3] David, Robin, et al. "BINSEC/SE : A dynamic symbolic execution toolkit for binary-level analysis." 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER). Vol. 1. IEEE, 2016.
- [4] Hart, Peter E., Nils J. Nilsson, and Bertram Raphael. "A formal basis for the heuristic determination of minimum cost paths." IEEE transactions on Systems Science and Cybernetics 4.2 (1968) : 100-107.
- [5] De Castro Pinto, Theo, et al. "Guiding Symbolic Execution with A-Star." International Conference on Software Engineering and Formal Methods. Cham : Springer Nature Switzerland, 2023.
- [6] "Welcome to the Wookey Project Documentation!" Welcome to the WooKey Project Documentation! - Wookey 0.9.0 Documentation, wookey-project.github.io/. Accessed 9 Apr. 2024.



# COMPUTER-ASSISTED ROUGH 2D ANIMATION INBETWEENING

Melvin EVEN

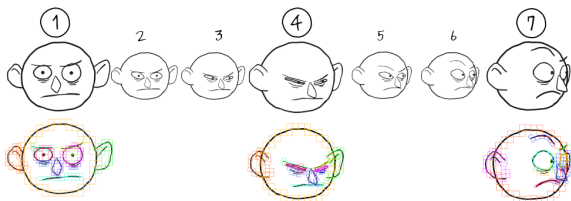
Direction : Pascal BARLA  
Co-Direction : Pierre BÉNARD

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

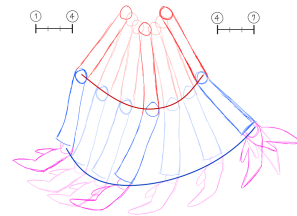
**T**raditional 2D animation requires a lot of planning, not only at the level of storyboards, but also for the animation itself where rough drawings (i.e. sketches) are used to test out the motion of different characters or special effects (e.g. water, smoke). I am interested in providing artists new ways to simplify this task and stylize animations. In the following, I briefly present three contributions : the first two about the creation and stylization of 2D motion and the last one about the style-preserving synthesis of sketchy drawings.

## Motion previsualization

Traditional 2D animation requires time and dedication since tens of thousands of frames need to be drawn by hand for a typical production. Many computer-assisted methods have been proposed to automatize the generation of inbetween frames from a set of clean line drawings, but they are all limited by a rigid workflow and a lack of artistic controls, which is in the most part due to the one-to-one stroke matching and interpolation problems they attempt to solve. We take a novel view on those problems by focusing on an earlier phase of the animation process that uses rough drawings (i.e. sketches). Our key idea is to recast the matching and interpolation problems so that they apply to transient embeddings, which are groups of strokes that only exist for a few keyframes. A transient embedding carries strokes between keyframes both forward and backward in time through a sequence of transformed lattices (Fig. 10.1, second row). Forward and backward strokes are then cross-faded using their thickness to yield rough inbetweens. With our approach, complex topological changes may be introduced while preserving visual motion continuity. As demonstrated on state-of-the-art 2D animation exercises, our system provides unprecedented artistic control through the non-linear exploration of movements and dynamics in real-time.



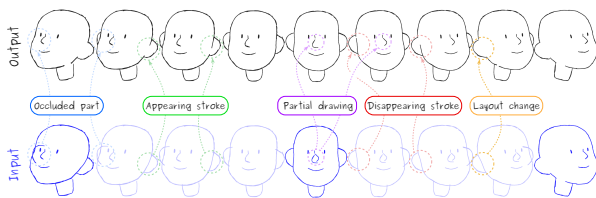
**FIGURE 10.1** – First row : given the input keyframes 1, 4 and 7, our system synthesizes inbetween frames in real-time. Second row : transient embeddings are shown at keyframes.



**FIGURE 10.2** – Inbetween frames can be further refined in both space and time via interactive tools based on traditional 2D animation : timing charts (top-left and top-right corners) as well as drawn trajectories (red and blue curves).

## 3D-like motion

Representing 3D motion and depth through 2D animated drawings is a notoriously difficult task, requiring time and expertise when done by hand. Artists must pay particular attention to occlusions and how they evolve through time, a tedious process. Computer-assisted inbetweening methods such as cut-out animation tools allow for such occlusions to be handled beforehand using a 2D rig, at the expense of flexibility and artistic expression. In this second work, we extend the more flexible 2D animation framework presented in the first section [1] to handle occlusions. We do so by retaining two key properties of their system : input rough drawings, and non-linear animation editing. Our contribution is two-fold : a fast method to compute 2D masks from rough drawings with a dynamic layout system for occlusions between parts ; and an artist-friendly method to control the dynamic visibility of strokes for self-occlusions. Our system helps artists produce convincing 3D-like 2D animations, including head turns, foreshortening effects, out-of-plane rotations, overlapping volumes and even transparency (Fig. 10.3).



**FIGURE 10.3** – Masks are automatically computed for all drawing parts to create occlusions among them; a stroke is appearing through a temporal visibility threshold; a partial drawing of the nose is created at an intermediate frame (it is stored at a keyframe, but is made to appear later on via visibility); a stroke is made to disappear on the opposite ear; and it is suddenly interrupted by the “pop” in front of the right ear via a layout change in between keyframes.

## Rough drawing interpolation

In this third work [2], we propose a novel efficient interpolation algorithm that generates intermediate frames in the artist’s drawing style. Starting from a set of registered rough vector drawings, we first generate a large number of candidate strokes during a pre-process, and then, at each intermediate frame, we select the subset of those that appropriately conveys the underlying interpolated motion, interpolates the stroke distributions of the key drawings, and introduces a minimum amount of temporal artifacts. In addition, we propose quantitative error metrics to objectively evaluate different stroke selection strategies. We demonstrate the potential of our method on various animations and drawing styles, and show its superiority over competing raster and vector-based methods.



**FIGURE 10.4** – Our interpolation technique takes as input a series of rough vector key drawings (shown in light gray), and produces intermediate drawings (in black) that preserve the input drawing style(s). It is based on an interactive stroke distribution synthesis algorithm that introduces a minimum amount of temporal artifacts

## Bibliographie

- [1] Even, M., Bénard, P. and Barla, P. (2023), Non-linear Rough 2D Animation using Transient Embeddings. *Computer Graphics Forum*, 42 : 411-425. <https://doi.org/10.1111/cgf.14771>
- [2] Chen, J., Zhu, X., Even, M., Basset, J., Bénard, P. and Barla, P. (2023), Efficient Interpolation of Rough Line Drawings. *Computer Graphics Forum*, 42 : e14946. <https://doi.org/10.1111/cgf.14946>

# MODELING THE APPEARANCE OF MATERIAL IRIDESCENCE BASED ON PERIODIC STRUCTURES

Gary FOURNEAU

Direction : Pascal BARLA

Co-Direction : Romain PACANOWSKI

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

In nature, many animals, plants and gemstones have an iridescent appearance. These iridescences are due to specific geometric structures at scales comparable to visible wavelengths, resulting in what are known as structural colors. The most striking examples are photonic crystals, where the same structure is repeated in one, two or three dimensions, increasing the scale and complexity of interference effects. In addition, natural materials are often subject to irregularities, such as misalignments, which generate non-negligible appearance characteristics.

## Bragg mirrors

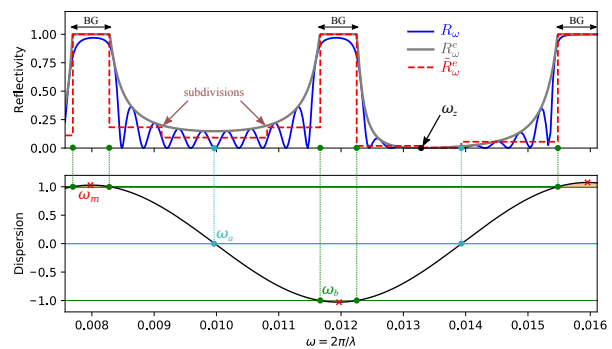
### Introduction

Various structures can generate iridescence, including thin films, diffraction gratings, and photonic crystals. Photonic crystals, in particular, are periodic dielectric structures that may exhibit repetitions in one, two, or three dimensions. Beginning with one-dimensional photonic crystals, also referred to as Bragg mirrors, they involve periodic arrangements of thin layers with differing optical indices. These structures are prevalent in biological organisms, notably in beetles, as depicted in Figure 11.2. However, simulating materials containing Bragg mirrors poses challenges due to their intricate spectra in both reflection and transmission. The aim is to develop rendering techniques facilitating interactive exploration of materials featuring Bragg mirrors.

### Simplify the reflectance spectrum

One initial challenge arises from the reflectance spectra of Bragg mirrors. As the number of periodic repetitions increases, high-frequency spectral oscillations begin to emerge. However, there are subsets of wavelengths known as photonic band gaps where reflectances approach values close to one, as illustrated in Figure 11.1. This phenomenon is primarily responsible for the vibrant reflectance colors observed. For simplifying the spectrum, we rely on the closed-form spectral reflectance formula proposed by Pochi Yeh [1]. In comparison to the standard transfer matrix approach, this method offers the advantage of being independent of the number of periodic repetitions. In addition to its performance benefits, this approach provides valuable insights into the structure of a Bragg mirror reflectance spectrum, particularly regarding the positioning of its band gaps. Leveraging these insights, we derive an approximate Bragg reflectance spectrum conducive to rapid, closed-form spectral integration for RGB rendering. This proves especially beneficial for swiftly exploring

the impact of various combinations of Bragg mirror parameters on the appearance of iridescent materials.



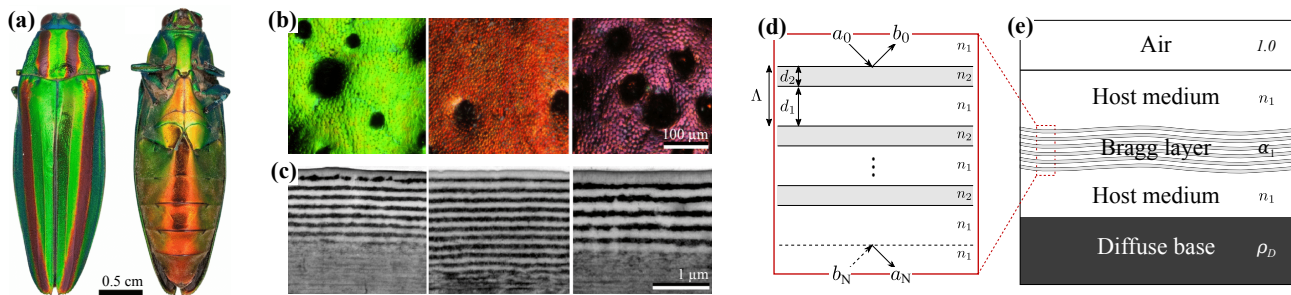
**FIGURE 11.1** – Top row : Reflectance spectrum  $R_\omega$  (in blue) of a Bragg mirror. The reflectance envelope  $R_\omega^e$  (in gray) equals 1 inside photonic band gaps (BG). Bottom row : the dispersion relation (in black) characterizes the location of a BG (it is either above 1 or below  $-1$ ). Spectral modes  $\omega_m$  (red crosses) always lie inside a BG and are used to find Airy points  $\omega_a$  (in cyan) and BG edges  $\omega_b$  (green dots). These spectral landmarks, along with zero points  $\omega_z$  (in black) are used to define a piecewise-constant spectrum  $\tilde{R}_\omega^e$  (dashed red curve in top row), which may be optionally subdivided.

### Model the roughness

Another challenge stems from the spatial arrangement of Bragg mirrors. Drawing inspiration from biological structures, we examine materials composed of one or more layers of Bragg mirrors arranged in different orientations atop a diffuse opaque base, as illustrated in Figure 11.2(d,e).

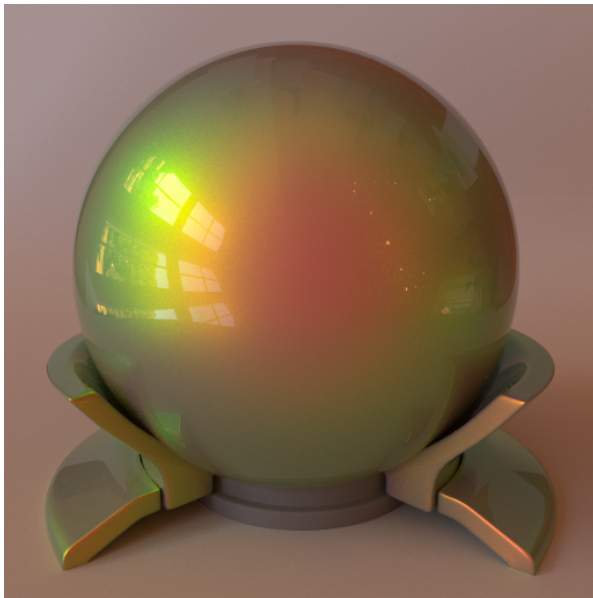
These Bragg layers are modeled using microfacet theory, where each microfacet represents a tiny Bragg mirror. Since only dielectric media are involved, a significant proportion of the incoming light reaching a Bragg layer is transmitted to layers below. We then use this property to derive a single-reflection model that further speeds up the rendering process, granting interactive feedback





**FIGURE 11.2** – 1D photonic crystals are commonly found in nature. Dorsal and ventral views (a) of a female Japanese jewel beetle, *Chrysochroa fulgidissima*. Microscopy (b) and TEM (c) images of its cuticle at different locations : the rough cuticle surface is shown in (b), while Bragg- like structures are observed in (c). All images from Schenck et al. [SWS13] - © IOP Publishing, reproduced with permission, all rights reserved. A Bragg mirror (d) is made of  $N$  cells, each composed of two thin films of different refractive indices ( $n_1, n_2$ ) and thicknesses ( $d_1, d_2$ ). The entry and exit refractive indices are equal, making the last interface index-matched (dashed line). A layered Bragg structure (e) consists of a rough Bragg layer (a contiguous distribution of Bragg mirrors of varying orientations) in a host medium of index  $n_1$ .

in spectral renderings. An example of a rendered result is depicted in Figure 11.3



**FIGURE 11.3** – Spectral rendering of a single rough Bragg layer on top of a diffuse base under a smooth coating. Parameters :  $n_1 = 1.4$ ,  $n_2 = 1.7$ ,  $d_1 = 114nm$ ,  $d_2 = 266nm$ ,  $N = 10$ . Albedo of the diffuse base :  $\rho_d = 0.2$

## Cholesteric Liquid Crystals

Cholesteric Liquid Crystals (CLCs) consist of rod-like nanostructures organized in a stable helical layered structure. In this case, a refractive index modulation arises from the birefringence of the nano-rods : its index varies depending on whether the incident light is polarized parallel or perpendicular to the rod. This periodic modulation also generates iridescent behavior. CLCs are abundantly present in natural structures such as in the pollia fruit and in certain beetles. They can also be easily replicated using various techniques documented in the literature. Numerous research works focus on manipulating the physical properties of CLC structures, including pitch, refractive index, number of repetitions, and birefringence, to modify their visual appearance.

The reflectance spectrum of CLCs shares similarities with that of Bragg mirrors. However the equations describing the behavior of CLCs differs from those of Bragg mirrors and we cannot use the exact same approach to simplify it. Currently, we are actively working on developing an efficient model to render this type of material effectively.

## Bibliographie

- [1] Pochi Yeh (1988). Optical waves in layered media. New York (N.Y.) : Wiley, 1988.

# BIOINFORMATIC METHODS TO STUDY METABOLISM AND CELL TYPES IN GLIOBLASTOMA

Johanna GALVIS

Direction : Macha NIKOLSKI, Bioinformatics and Computational Biology team

Co-Direction : Thomas DAUBON, Glioblastoma metabolism team

Univ. Bordeaux, CNRS, Institut de Biochimie et Génétique Cellulaires IBGC, UMR 5095, Bordeaux, France

The development of bioinformatic tools for multi-omics data analysis with a focus into metabolism, data integration, and tumoral microenvironment, is covered in the presented PhD project. The biological context is the Glioblastoma (GB), a lethal brain tumor. The first part of this project resulted in a new bioinformatic tool, DIMet, which performs the differential analysis of labeled metabolomics data, and its integration to transcriptomics data. DIMet was showcased on the blocked lactate transporter GB datasets, and accepted for publication (in press). During the second part, the implementation of methods to analyze spatially resolved metabolomics data is ongoing, and a fourth year is required.

The Glioblastoma (GB) metabolic plasticity plays a role in the invasiveness of the malignant cells, which is studied through the lens of high-throughput omics data [1]. The present interdisciplinary doctoral project, currently in its third year, is dedicated to the development of bioinformatic tools for the omics data analysis in this biological context. It is divided into two parts :

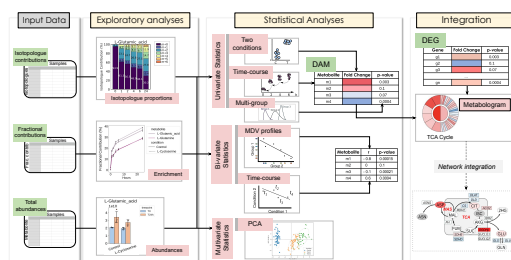
## Bulk labeled metabolomics data : analysis and multi-omics integration

**Rationale.** The labeled metabolomics technology allows to quantify the cells' metabolic activity, using a  $^{13}\text{C}$  labeled substrate to trace the speed of the production/-consumption of the metabolites. Whether with LC-MS, GC-MS or NMR<sup>2</sup>, the  $m/z$  peaks of isotopologues<sup>3</sup> are measured, then transformed to arrays of corrected intensities stored as tabular files. Available bioinformatic methods to compare labeled metabolomics data across conditions or time-points, have limitations : the fluxes quantification is expensive in computational resources, and assumes priors<sup>4</sup>. Other tools lack a rigorous statistical framework, or do not take into account the labeling patterns, commonly represented as mass distribution vectors (MDVs) [2].

**Objectives.** The objectives of this first part of the doctoral project are : (i) to automatize the differential analysis of labeled metabolomics data, and (ii) to develop a metabolic-pathway based strategy to integrate metabolomics and transcriptomics datasets<sup>5</sup>.

**Datasets.** In the GB P3 cell line, the knock out (KO) of the lactate transporter LDHA/B was associated with a

decrease in invasion *in vitro* [1]. Separately, for LDHA/B KO and Control samples, both labeled metabolomics and transcriptomics data were acquired, in *bulk* (i.e. no spatial nor single-cell resolution), and in triplicates. All the datasets are available at Zenodo.



**FIGURE 12.1** – DIMet accepts labeled metabolomics metrics : isotopologues, mean enrichment, and total abundances (left). Statistical analyses (center-right) explained in the Methods : the Fold Changes and  $p$ -values of the differentially abundant or labeled metabolites (DAM) are computed by DIMet. The integration (right panel) with the transcriptome (differentially expressed genes (DEG) obtained with standard analysis using e.g. DESeq2) through metabolograms is available. Metabolograms can be projected onto a network (by external tools).

**Methods.** The tool was developed in Python 3.9.7 and 3.10, its architecture is powered by the Hydra library, a versatile multi-configuration manager. Univariate statistical analysis (Figure 12.1) can be performed with non-parametric tests, whether comparing conditions or as a time course analysis. The bivariate analysis allows to compare the MDV profiles using the Spearman or Pearson correlation test. The  $p$ -values can be adjusted for multiple comparisons. The multivariate analysis is

2. Liquid Chromatography coupled to Mass Spectrometry, Gas Chromatography-MS, and Nuclear Magnetic Resonance, respectively

3. An isotopologue is a unique variant of the metabolite with a specific number of stable isotopes and mass, function of the number of labeled carbon atoms.

4. Flux computation relies on priors such as stoichiometric constants and a fixed repertoire of reactions.

5. Both omics being complementary : transcriptomics measures the expression of genes encoding for enzymes catalyzing the metabolic reactions

possible with the Principal Components Analysis. Visualizations can be customized. Moreover, DIMet integrates transcriptomics and labeled metabolomics data through pathway-based metabolograms. All above relies in SciPy, NumPy, Statsmodels, Pandas, Matplotlib and Seaborn libraries, among others.

**Results.** We present to the scientific community the new, open-source, one-stop comprehensive tool for differential analysis of targeted labeled metabolomics data : DIMet. It supports differential comparison in multiple modalities (e.g. time-course analysis, or entire MDV profiles). The use of DIMet in real datasets from GB P3 cell-line samples was illustrated. The differential results in labeling can be interpreted as a proxy to the reactions speed. DIMet is available at <https://github.com/cbib/DIMet>. For non-developers, a user friendly interface is accessible through <https://usegalaxy.eu>, and is also deployed in the Galaxy Workflow4Metabolomics. In conclusion, the two aforementioned objectives were met, and the associated paper has been accepted for publication in the *Application Notes* section of the *Bioinformatics* journal.

As an aside note, the development of a more advanced strategy for omics integration through interactive and editable metabolic networks was also explored. The graph operations were defined, a DSL editing language was proposed, and the automatic selection of sub-networks was set. After initial front-end experimentation, serious limitations concerning the graph layout were found. Unfortunately, due to time constraints, this work was suspended. Indeed the second part of the doctoral project needs more urgent attention.

### Spatial Omics data analysis to study the microenvironment's metabolic state

The second part of the project is dedicated to the development and automation of methods for the Spatial Metabolomics data analysis. The Spatial Metabolomics data, in our context, consists of the metabolic profiles measured in *ex-vivo* brain tissue slices of Xenografted human GB cells, in a murine model. Under the hypotheses that : a) the metabolic state of the GB microenvironment is the result of metabolic interactions between the different cell types, and b) that the Spatial Metabolomics data captures cell-type signals, the 2 following objectives were fixed :

1. To computationally establish the composition of the 2-dimensional metabolomics data, in relation to the cell types, namely, astrocytes, oligodendrocytes, neurons, and two types of tumoral cells.
2. To propose a representation of the metabolic state of the microenvironment taking into account the cell types and their relative location in the space.

**Datasets.** Briefly, the tissue slice is subjected to a matrix fixation, ion preparation and compounds detection with a MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization coupled to a Time-of-Flight mea-

surement device) instrument. For each acquisition, the spectra of metabolites' mass/charge ( $m/z$ ) ratios (peaks), across all the pixels of the slice, are generated. These spatially resolved spectra, known as Mass Spectrometry Imaging (MSI) data, are converted to arrays of intensities (imzML format). To note, peaks are ambiguously mapped to compounds identifiers of the Human Metabolism Data Base (HMDB) : one peak often corresponds to more than one identifier. These HMDB annotations are registered into tabular files for a later use, for instance, to perform pathway mappings using the Kyoto Encyclopedia of Genes and Genomes(KEGG). A cooperation with the PRISM -*Protéomique Réponse Inflammatoire Spéctrométrie de Masse*- laboratory at Lille University was established, providing us the raw imzML files for our bioinformatic developments at the IBGC. The received MALDI-TOF MSI data corresponding to two acquisitions (LDHA/B KO and Control) have a pixel size of  $50\mu m$ , an average of 40000 pixels, and a  $m/z$  range between 88 and 200.

**Methods.** Each imzML file has been converted to a 2-dimensional dense matrix of positive floats, where the columns correspond to the pixel coordinates, and the rows correspond to ( $m/z$ ) peaks. Different non-supervised machine learning strategies for the attribution of pixels to discriminative groups were explored : K-Nearest Neighbors (KNN),  $k$ -means, hierarchical clustering. However, preliminary results show that these strategies tend to group the pixels by patterns of combined signals. The Linear Discriminant Analysis (LDA) performs multilabel classification, but requires adequate training datasets. A probabilistic cell-type composition of the pixels via supervised classification algorithms is recommended. Unfortunately, ground-truth datasets are scarce or not centralized, and fingerprints of purified cell types are not yet available (TD is considering to acquire them shortly). In view of these challenges, we are moving towards the development of network representations of cell type metabolic profiles, and the use of Graph Neural Network algorithms. Therefore, this part of the project will be extended into a 4th year of Thesis.

### Bibliographie

- [1] Guyon,J. et al. (2022) Lactate dehydrogenases promote glioblastoma growth and invasion via a metabolic symbiosis. *EMBO Mol Med*, 14.
- [2] Buescher,J.M. et al. (2015) A roadmap for interpreting  $^{13}C$  metabolite labeling patterns from cells. *Current Opinion in Biotechnology*, 34, 189–201.

# EXPLORING VIRTUAL REALITY BY AND FOR NON-EXPERT USERS

Edwige CHAUVERGNE

Direction : Martin HACHET

Co-Direction : Arnaud PROUZEAU

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

If virtual reality (VR) was initially only available to experts, it has nowadays been introduced to the general public. VR headsets are being used in exhibitions, entertainment centers and at home. Yet, introducing new technology to the general public comes with challenges. During this thesis, we focused on three main challenges, 1) learning how to use the hardware and new immersive applications, 2) offering some immersive content suited for non-experts, and 3) empowering non-experts through enabling the authoring of VR content by non-experts for non-experts.

## User Onboarding in Virtual Reality

Using a new software always comes with a learning curve whether the application is in VR or on more traditional hardware such as a computer. We call this learning phase *onboarding*, which is defined by Renz et al [1] as "the sum of methods and elements helping a new user to become familiar with a digital product".

We investigated the current practices and challenges in VR onboarding [3]. To do so, we reviewed 21 tutorials based on a heuristic evaluation and we interviewed 15 VR experts with experience in VR onboarding.

We gathered the results of our tutorial review into three main challenges : 1) Intelligibility, 2) Feedback and feedforward, and 3) Adaptability and accessibility.

| Onboarding Context      |                     |                                                      |                                              |
|-------------------------|---------------------|------------------------------------------------------|----------------------------------------------|
| Out of context          |                     | Formation specific context                           | Context-aware                                |
| Degree of flexibility   |                     |                                                      |                                              |
| Linear and rigid        | Skippable and rigid | Skippable and flexible                               | Focus on specific functionalities            |
| Degree of coverage      |                     |                                                      |                                              |
| Basic functionalities   |                     | All the functionalities without level discrimination | Level specific training                      |
| Inputs onboarding       |                     |                                                      |                                              |
| Outside VR              |                     | With visual cues                                     | No onboarding                                |
| Degree of interactivity |                     |                                                      |                                              |
| Observer                |                     | Partially enabled functionalities                    | Freely interact with the virtual environment |
| Type of assistance      |                     |                                                      |                                              |
| Instructions            |                     | Demonstration                                        |                                              |
| Degree of feedback      |                     |                                                      |                                              |
| Implicit feedback       |                     | Explicit feedback                                    | Corrective feedback                          |
| Instruction modalities  |                     |                                                      |                                              |
| Visual                  |                     | Sound                                                | Haptic                                       |

FIGURE 13.1 – Conceptual framework for general assisted VR onboarding

We did a thematic analysis [2] of our interviews and identified three main challenges : 1) the novelty of VR, 2) the lack of awareness, and 3) the communication barrier in VR.

Based on our tutorial review and our expert interviews we created a framework for general onboarding and provided guidelines for general and instructor-based onboarding. Our framework comprises eight dimensions and is depicted in figure 13.1. We expect this framework to become a support to researchers and onboarding designers as (1) it has descriptive power, (2) it can be used to compare designs, and (3) it supports generating new designs. However, this framework is not meant to be prescriptive, as further research is required to understand which design better fits a situation.

## Visualizing humanitarian data in virtual reality

As VR headsets become available to the general public, it offers new opportunities for content diversification and it would be interesting to investigate the creation of accessible content for non-experts in VR. In our case, we focused on data visualization. In some cases, the use of VR can facilitate the analysis of complex data by experts thanks to the immersive and three-dimensional environment it offers. However, when introducing data to the general public, our main interest in VR might not be because of its capacity to display data in 3D, but rather because of the sense of immersion and reality it can provide.

VR can be useful when communicating about humanitarian data. It is indeed a promising tool to raise empathy thanks to the sense of presence and realism it can provide. Yet, immersive content meant to raise empathy such as VR documentaries, usually gives very little to no information regarding the extent of a crisis for a quantitative understanding. Thus, we inspired ourselves with the concept of semantic zoom and designed a new type of visualization in VR that transitions from an abstract traditional visualization, to an immersive embodied visualization.

We explored the potential of this type of visualization



through three use cases : 1) the presumed femicides in France in 2022 registered by the collective "Féminicides par compagnons ou ex"<sup>6</sup> (See figure 13.2)), 2) the bike accidents that occurred in Gironde in 2018<sup>7</sup> and, 3) and exploration of animal welfare through the visualization of the laying hens condition. The three use cases demonstrate that beyond showing the human (/living) being behind the data, the immersive view can be used to provide context, embody the victim and inform on non-codable data.



FIGURE 13.2 – The immersive views of the visualization of the femicides in France in 2022.

## Authoring Interactive and Immersive Experiences Using Programming by Demonstration

The last step to introduce and settle the use of VR for the general public is to empower the users by enabling the prototyping and authoring of immersive VR experiences, which allows non-developers to appropriate the use of VR and create new content.

Various methods have been investigated to enable development for non-experts such as simplified desktop interfaces, visual programming [4], and authoring by demonstration [5]. However, these approaches do not handle yet the creation of complex interactive experiences.

Immersive authoring by demonstration appears as a promising creation process for end-users as its users directly manipulate objects in context and can immediately visualize and experience the created interactions. In order to better understand the challenges to the immersive authoring of interactive experiences by demonstration, we implemented an authoring tool, created two use cases, and organized a feedback session [6].

We implemented a prototype in which users can record actions. As they start the recording, they can directly manipulate objects in the virtual environment to demonstrate actions. After recording actions, users can link actions as triggers and effects. Finally, users can directly test the interactions they authored by activating a test mode.

Based on the use cases' observations and the feedback sessions' results we identified five main challenges for

authoring interactive experiences by demonstration : 1) Handling impossible demonstrations, 2) Demonstration of abstract concepts, 3) Processing ambiguities, 4) Creating and visualizing logic and, 5) Providing assistance to users.

## Conclusion

Making VR headsets affordable and broadly available is only the first step to supporting the use of VR by the general public. More research required to support the onboarding and authoring of interactive experiences and the benefits of VR for end-users still need to be explored. We hope this PhD will provide some guidelines to reach this goal.

## Bibliographie

- [1] Renz, J., Staubitz, T., Pollak, J., and Meinel, C. (2014). Improving the onboarding user experience in MOOCs. In EDULEARN14 proceedings (pp. 3931-3941). IATED.
- [2] Virginia Braun and, Victoria Clarke (2006) Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3 :2, 77-101.
- [3] Chauvergne, E., Hachet, M., and Prouzeau, A. (2023, April). User onboarding in virtual reality : An investigation of current practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).
- [4] Lei Zhang and Steve Oney. 2020. FlowMatic : An Immersive Authoring Tool for Creating Interactive Scenes in Virtual Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 342-353.
- [5] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Yuanzhi Cao, and Karthik Ramani. 2021. GesturAR : An Authoring System for Creating Freehand Interactive Augmented Reality Applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 552-567.
- [6] Chauvergne, E., Hachet, M., and Prouzeau, A. (2023, April). Authoring Interactive and Immersive Experiences Using Programming by Demonstration. In *Proceedings of the 34th Conference on l'Interaction Humain-Machine* (pp. 1-13).

6. [www.feminicides.fr](http://www.feminicides.fr)

7. <https://www.data.gouv.fr/fr/datasets/accidents-de-velo/>

# MODELING THE ECO-EVOLUTIONARY ORIGINS OF OPEN-ENDED SKILL ACQUISITION

Gautier HAMON

Direction : Clément MOULIN-FRIER

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

Humans have evolved the ability to learn new skills to adapt to a variety of situations they might encounter. This ability is of great interest in the study of artificial intelligence, which also aims to develop this level of adaptation and learning. In fact, instead of trying to learn a specific task, the field of artificial intelligence is interested in learning/evolving agents that are able to adapt and learn quickly new things : ie learning to learn or "meta learning". In particular, as learning a skill might be a stepping stone to learn new more complex skills, an agent that would constantly learn and accumulate knowledge could be able to learn always more complex skill, what we call "open-ended skill acquisition". In this thesis, we aim to study what are the ecological condition that would give rise to the emergence of open-ended skill acquisition? Studying the impact of environmental features on the learning process. We study it both at a low level of cognition where we look at how evolution might be able to evolve simple entities into more and more complex ones. And at a higher level where agent at a developmental scale (ie during their lifetime) face uncertain or variable environments and have to adapt to it quickly through learning from experience and exploration.

## Towards open ended evolution in cellular automata

In these work [1, 2], we position ourselves at a low level of cognition to try to build a system that would reproduce some kind of open ended evolution. Meaning that the system would continuously evolve more and more complex agents. And in particular we are interested in studying the minimal environmental conditions that would give rise to this open-ended evolution.

In a first project [1], we first looked at the existence of cellular automata (CA) rules that would self-organize agent with some very basic cognitive ability : sensorimotor capabilities, ie the ability to react to perturbation from the environment. This project included the introduction of obstacles as environmental perturbation in the system. We then used machine learning techniques such as curriculum learning, gradient descent, diversity search to optimize the rule of the CA in order to self-organize those sensorimotor agents.

In a second project [2], we extend a known cellular automaton Lenia with mass conservation, introducing flow Lenia. This allows to add environmental constraint more easily as well as competition for mass and evolution. In addition, this new CA allows to have local parameters, allowing to have "multi-species" simulation (ie simulation with several cellular automata rules on different part of the grid). Such a feature opens up research perspectives towards the achievement of open-ended intrinsic evolution inside continuous CAs (see Fig.14.1), which means that all the evolutionary part would be a result of the dynamic of the CA (without any external loop/system). We hypothesize that this open-

ended intrinsic evolution could, through the competition/cooperation, lead to the emergence of interesting low level cognition in those system.

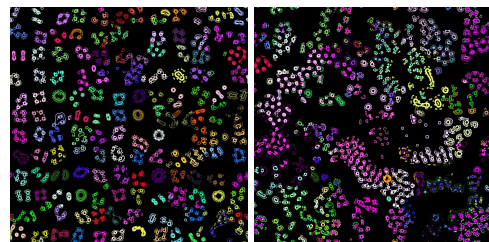


FIGURE 14.1 – Intrinsic evolution in a multi species simulation in Flow Lenia

## Eco-evolutionary dynamics

Environment plays an important role in open-ended evolution, but in most evolutionary algorithms works, the environment experience reset at each generation. This removes an important part of evolution where "organisms are not solely products but, by modifying their niche and therefore its associated fitness landscape, are also causes of evolution" [3]. The main objective of the paper[4] is to propose a simple large-scale eco-evolutionary dynamics in agent-based simulations with a reasonable level of biological and ecological plausibility and measure its dynamic. The environment contains a variable population of agents (often more than a hundred), each controlled by a neural network whose weights are optimized using neuroevolution.

We prevent any environment or population reset during a whole simulation run, enabling coupled environmental and population dynamics leading to complex



eco-evolutionary feedback effects.

## Multi agent Autotelic agents

The intrinsically-motivated goal-conditioned learning paradigm is well-established in single-agent settings : by setting its own goals and pursuing them in an environment without external supervision, such an agent is able to acquire a wide diversity of skills [5]. Such agents are called autotelic, from the Greek words auto (self) and telos (end). This framework given a way to sample goals and to assess their success lead to agent acquiring more and more complex skills.

However, What happens when you transfer the autotelic paradigm to multi-agent environments, where some skills may require the cooperation of multiple agents (lifting a heavy box for example) ? This is the question we aimed at addressing in this project [6]. We believe that multi-agent applications will benefit from agents that can autonomously discover and learn cooperative skills, but entail additional challenges to the ones found in single-agent settings : agents that independently set their own goals will have a very low probability of simultaneously sampling the same cooperative goal, which will make solving these goals difficult. We address this issue by allowing the agent to send messages to the other before choosing a goal.(though the agent have to learn which message to send and coordinate with the other)

## Meta learning : Learning to learn

Instead of relying on the explicit sampling of goal which often has to be predefined. Training over a distribution of tasks can allow to meta learn this sampling of goal end to end in the dynamic of the agent. For example, training on a distribution of environment, where for each environment the agent doesn't know the rules and can only learn them through trials and errors, should lead the agent to meta learn an efficient exploration strategy to learn as quickly as possible the rule of the environment to exploit them.

In particular, as a first step towards studying the evolution of open-ended skill acquisition in artificial agents, we studied the environmental conditions favoring the systematic exploration of combinatorial recipes involving various objects. By combinatorial recipe, we mean the ability of agents to combine objects in the environment in order to create new ones (in the spirit of the Minecraft video game). In this work, the training of an agent uses meta reinforcement learning where an outer loop, equivalent to an evolutionary mechanism, meta-learns the parameters of an inner loop which can be seen as a developmental mechanism (where the agent acquire skills during its lifetime by interacting with the environment). In particular in this setup the agent effectively learns to explore and memorize recipe during its lifetime.

B) Example of episode rollout

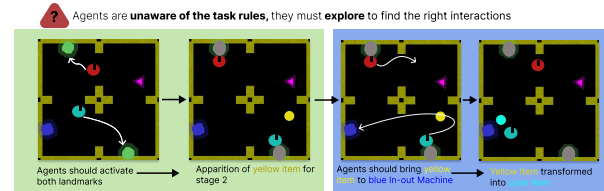


FIGURE 14.2 – Example of task environment.

In another work [7], we further investigate the emergence of cooperative exploration strategies of decentralized agents by training them on a open ended distribution of tasks. To this end we introduce a novel environment which is conceptually simple yet allows for a complex open ended procedurally generated task space by dynamically combining multiple subtasks sampled from five task types to form a task tree which needs to be solved sequentially, akin to the notion of recipes (See fig 14.2 for an example of task). As no information is given to the agents about which subtasks have been sampled or how and in which order they should be solved, the agents have to develop general strategies for exploring the environment, effectively learning how to learn from the information obtained by interacting with the environment throughout the episode, in order to solve novel tasks.

## Bibliographie

- [1] Hamon.G et al. (2024). Discovering Sensorimotor Agency in Cellular Automata using Diversity Search. arXiv preprint arXiv :2402.10236
- [2] Plantec.E et al ; 2023. "Flow-Lenia : Towards open-ended evolution in cellular automata through mass conservation and parameter localization." Proceedings of the ALIFE 2023
- [3] Laland Kevin N. et al. 2015.The extended evolutionary synthesis : its structure, assumptions and predictionsProc. R. Soc. B.28220151019
- [4] Hamon.G et al. 2023. Eco-evolutionary Dynamics of Non-episodic Neuroevolution in Large Multi-agent Environments. In Proceedings of the Companion Conference on Genetic and Evolutionary Computation (GECCO '23 Companion)
- [5] Colas.C et al. 2022. Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning : a short survey. Journal of Artificial Intelligence Research v-74.
- [6] Nisioti.E et al.2023. Autotelic Reinforcement Learning in Multi-Agent Environments ; Proceedings of The 2nd Conference on Lifelong Learning Agents, PMLR 232 :137-161
- [7] Bornemann, R et al. (2023). Emergence of Collective Open-Ended Exploration from Decentralized Meta-Reinforcement Learning. arXiv preprint arXiv :2311.00651.

# ALGORITHMES DE COUVERTURE POUR LES RÉSEAUX DE PETRI CERTIFIÉS EN COQ

**Thibault HILAIRE**

**Direction : Jérôme LEROUX**

**Co-Direction : David ILCINKAS**

*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

**L**es systèmes concurrents sont des systèmes qui peuvent utiliser plusieurs processus de façon simultanés qui vont se partager des ressources communes. Ils sont de plus en plus présents dans nos systèmes informatiques modernes et en particulier dans des systèmes critiques.

Cependant leur correction, c'est-à-dire l'absence de bugs et autres failles, n'est pas chose facile due au fait qu'ils doivent fonctionner correctement qu'importe l'ordre d'exécution des différents processus.

C'est pour cela que dans ma thèse je m'intéresse à la vérification formelle, et plus précisément au model-checking. Le but est d'abstraire des problèmes sur des systèmes complexes et de les représenter par des problèmes sur des modèles mathématiques plus simples. Il faut trouver des modèles qui soient suffisamment puissants pour abstraire beaucoup de systèmes complexes, mais qui ne soient pas trop puissants au point de rendre les problèmes sur ces modèles indécidables. Par si un modèle peut modéliser les machines de Turing alors beaucoup de problèmes seront indécidables et donc le modèle inutilisable.

Les réseaux de Petri constituent un modèle bien étudié pour vérifier les systèmes concurrents. Le calcul de l'ensemble de couvrabilité est l'un des problèmes les plus fondamentaux concernant les réseaux de Petri, et c'est celui qui m'intéresse dans le cadre de ma thèse.

Obtenir des preuves certifiées par machine serait donc un véritable gain vu la complexité d'écrire à la main des preuves complètes et correctes pour des systèmes concurrents. C'est pour cela que je me suis intéressé à l'assistant de preuve Coq. Avec ce dernier nous avons certifié la correction et la terminaison de l'algorithme MINCov de Finkel, Haddad et Khmelnitsky (FOSSACS 2020). Cet algorithme est l'algorithme le plus récent dans la littérature qui calcule la base minimale de l'ensemble de couvrabilité, un problème connu pour être sujet à des bugs subtils.

Outre l'intérêt intrinsèque d'une preuve vérifiée par ordinateur, notre certification fournit de nouvelles informations sur l'algorithme MINCov. En particulier, nous introduisons comme algorithme intermédiaire une variante à petites étapes de MINCov d'un intérêt indépendant.

## Techniques et méthodes

### Coq

Coq est un assistant de preuve, c'est-à-dire un logiciel permettant la vérification de preuves mathématiques comme des théorèmes ou encore des assertions relatives à l'exécution de programmes informatiques. Attention, cela reste un assistant, il ne va pas faire vos preuves à votre place. C'est un logiciel libre développé par l'équipe PI.R2 de l'Inria et en partenariat avec l'école polytechnique, le CNAM, l'Université Paris Diderot et l'Université Paris-Sud.

Coq est fondé sur le calcul des constructions, une extension de la correspondance de Curry-Howard utilisant la logique propositionnelle intuitionniste. Le calcul des constructions utilisé dans Coq comprend directement les constructions inductives, d'où son nom de calcul des constructions inductives.

Le passage de la logique classique au calcul des constructions nécessite un temps d'adaptation. Effectivement dans nos preuves habituelles en informatique et en mathématiques, il n'est pas rare d'utiliser le tiers exclu ou

encore des preuves par l'absurde.



Bref des preuves qui montrent qu'un objet existe sans pour autant montrer comment le construire. Or Coq n'a pas naturellement ces principes puisqu'ils font partie de la logique classique. Ce qui nous force, si on décide de ne pas les rajouter, à changer nos preuves pour avoir des preuves constructives et directes. Cependant ces preuves ont le bon goût d'apporter plus d'informations sur ce qui se passe réellement lors de la preuve. Ce qui est d'autant plus vraie par le fait que Coq ne peut pas faire des bonds de raisonnement logique comme les humains, ce qui nous force lors de preuves en Coq à

vraiment décomposé nos preuves en briques élémentaires. Il y a donc aussi un côté pédagogique à faire des preuves en Coq.

Tout cela n'a pas empêché de montrer des gros résultats en Coq, notamment :

- Le théorème des quatre couleurs par Georges Gonthier et Benjamin Werner [1].
- Le théorème de Feit-Thompson par Georges Gonthier et son équipe [2].
- CompCert, un compilateur optimisant pour le langage C qui est en grande partie programmé et prouvé en Coq

Coq a déjà été utilisé pour vérifier l'algorithme de couverture le plus fondamental sur les réseaux de Petri dans [4]. Cette certification utilise la bibliothèque `MATHCOMP`[3] qui a été développé afin de pouvoir faire des preuves mathématiques formelles de haut niveau en Coq. Elle a été développée par Georges Gonthier et son équipe lorsqu'ils ont prouvé le théorème de Feit-Thompson.

J'ai donc étendu la modélisation sur les réseaux de Petri afin de pouvoir vérifier la terminaison et la correction de l'algorithme de couverture le plus récent, `MINCOV`. J'ai aussi proposé une variante à petits pas de cet algorithme, appelé `ABSTRACTMINCOV`, que j'ai aussi certifié avec Coq.

|           |                                                   |            |
|-----------|---------------------------------------------------|------------|
| [4]       | Technical tools                                   | 631 lines  |
|           | Petri net                                         | 1226 lines |
|           | Karp-Miller                                       | 775 lines  |
| Our paper | Technical tools                                   | 1780 lines |
|           | Petri net extension                               | 1873 lines |
|           | <code>MINCOV</code> & <code>ABSTRACTMINCOV</code> | 5634 lines |

Le code est accessible sur Software Heritage au lien suivant :

<https://archive.softwareheritage.org/swh:1:rev:7b5523e30026266c471c73e911f0fda525c6f900;origin=https://gitub.u-bordeaux.fr/thhilaire/petri-net-in-coq.git>.

## Bibliographie

- [1] Gonthier, G. The Four Colour Theorem : Engineering of a Formal Proof. *Computer Mathematics, 8th Asian Symposium, ASCM 2007, Singapore, December 15-17, 2007. Revised And Invited Papers.* **5081** pp. 333 (2007), <https://doi.org/10.1007/978-3-540-87827-8>
- [2] Gonthier, G., Asperti, A., Avigad, J., Bertot, Y., Cohen, C., Garillot, F., Roux, S., Mahboubi, A., O'Connor, R., Biha, S., Pasca, I., Rideau, L., Solovyev, A., Tassi, E. & Théry, L. A Machine-Checked Proof of the Odd Order Theorem. *Interactive Theorem Proving - 4th International Conference, ITP 2013, Rennes, France, July 22-26, 2013. Proceedings.* **7998** pp. 163-179 (2013), <https://doi.org/10.1007/978-3-642-39634-2>

- [3] Gonthier, G., Mahboubi, A., Tassi, E. : A small scale reflection extension for the Coq system. Ph.D. thesis, Inria Saclay Ile de France (2016)
- [4] Mitsuharu Yamamoto, Shogo Sekine, and Saki Matsumoto. Formalization of Karp-Miller tree construction on petri nets. In Yves Bertot and Viktor Vafeiadis, editors, *Proceedings of the 6th ACM SIGPLAN Conference on Certified Programs and Proofs, CPP 2017, Paris, France, January 16-17, 2017*, pages 66-78. ACM, 2017.

# PRACTICAL WHITE-BOX CRYPTOGRAPHY

Agathe HOUZELOT

Direction : Arnaud CASTEIGTS

Encadrement : Christophe GIRAUD and Emmanuelle DOTTA

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

IDEMIA, Cryptography & Security Labs, 33600 Pessac, France

---

**W**hite-box cryptography aims at protecting software implementations of cryptographic algorithms against a very powerful attacker who may gain control over the execution environment. If nowadays all published white-box designs for standard cryptosystems are considered broken, some solutions can actually be usable in practice and white-box implementations are currently sold by several companies. Indeed, the capacities of the adversary vary widely depending on the use case and some attacks that were introduced in the literature can actually be difficult to apply in practice. It is thus important to identify attacks and countermeasures that are relevant in real-life scenarios, and this is what I tried to do in my PhD thesis for two well-known schemes, namely AES and ECDSA.

---

## White-Box Cryptography

### Context and Brief History

The goal of white-box cryptography (WBC) is to protect the secret keys embedded in software implementations of cryptographic algorithms that are executed in untrusted environments. Such implementations require special care since cryptosystems are mainly designed to resist an attacker who can only observe inputs and outputs, and of course, an adversary controlling the execution environment of a software implementation could obtain far more information. In the white-box model, we consider that the attacker can analyze (or even alter) both the binary and the values read/written in memory.

The first white-box implementations were introduced by Chow et al. in 2002 [1, 2]. Since then, the field has received considerable attention from the community with many design proposals, attacks and efforts to define security notions. Unfortunately, to this day, all public implementations of standard cryptosystems have been broken and the question of the (im)possibility of protecting such algorithms in the white-box model remains open. However, some ad-hoc cryptosystems were specifically designed for this context and proved to satisfy some useful security notions.

### White-Box Cryptography In Practice

For reasons of interoperability, the industry is unlikely to adopt ad-hoc cryptosystems. Nevertheless, as already mentioned, all public designs of standard algorithms are currently broken. In such circumstances, the industry is compelled to devise proprietary solutions that rely primarily on obscurity, i.e. on the secrecy surrounding their implementation. The objective is to provide practical security by preventing automated attacks and applying code obfuscation to ensure that the adversary

has to go through a time-consuming and arduous reverse engineering phase.

While these practical solutions are less secure than what is expected in the white-box model, real-life adversaries may also be far less powerful than described. Indeed, the capabilities of attackers vary widely depending on the use case. For example, if the white box is bound to a mobile-payment application that implements authentication mechanisms, then the attacker should not be able to execute the code himself, rendering chosen plaintext attacks impractical and limiting the number of executions within a given time frame.

Practical white-box designs should thus be constructed for a specific purpose and come with a detailed adversary model so that relevant analysis can be done. Similarly, when a new attack is discovered, its practicality should be evaluated in terms of several parameters, including the possibility of automation, the required number of executions and the need to choose specific inputs.

## Automated Attacks

Automated white-box attacks are particularly interesting since they do not require any reverse engineering step. They encompass two well-known attacks that have already been extensively studied in the smart-card world : side-channel and fault attacks.

### Side-Channel Attacks

In 2016, Bos et al. [3] introduced the first automated attack against white-box implementations. The principle is very similar to the classical side-channel attacks (SCA) [4] that were originally designed to break hardware implementations : secret values are extracted from leakage traces obtained during several executions of a cryptographic algorithm with the help of statistical tools. The only difference relies upon the nature of the traces.



When attacking a hardware implementation, one would for instance measure the power consumption of the device during the execution of the algorithm. Instead of relying on such noisy information on the processed data, a white-box attacker can directly record traces of all the values that are read or written in memory, reducing a lot the number of required executions and leading to efficient practical attacks.

### **Fault Attacks**

Another automated attack consists in purposefully disturbing the algorithm execution and exploiting the resulting faulty output. It was first introduced in [5] against hardware implementations, but in the white-box context, errors can be induced even more easily since the attacker can modify the binary or use debugging tools to stop the execution and, for example, skip an instruction or modify the value of a register. Fault attacks are very powerful since they can be fully automated and usually only require a very limited number of executions on random plaintexts.

## **Main Contributions**

When my PhD started, almost all the work on WBC focused on the Advanced Encryption Standard (AES). I continued the research on that scheme but also got interested in the Elliptic Curve Digital Signature Algorithm (ECDSA). Indeed, white-box implementations of the latter are actually needed and already sold by the industry, which is in contrast with the lack of published articles on the subject. During my PhD, I contributed to a few publications that I describe below. I also co-authored a popular science article on WBC for *Interstices* and helped design and implement a proprietary ECDSA white box for my company.

### **Patent Study**

Our first paper [6] provides a deep comprehension on the challenges of ECDSA white-box implementations. It highlights the existence of particularly devastating attacks and gives an insight into the actual strategies for securing products in the industry. Given the lack of articles on ECDSA white boxes, we analyzed the only source of information that we had - which is the patents filled by companies - and discussed how they respond to existing threats.

### **Fault Attack on an ECDSA White Box**

The only white-box design for ECDSA that was published in the literature makes the strong assumption that a semi-honest cloud server can be accessed during the computation of a signature. In [7], we presented two very efficient fault attacks on that scheme, showing along the way that even with the help of a server, securing ECDSA in the white-box model really is challenging. We also proposed a lightweight countermeasure that prevents both our attacks.

### **Participation to the WhibOx Contest**

In 2021, the WhibOx contest was organized for the third time as part of the CHES workshops. Developers were invited to submit ECDSA white-box implementations and attackers to try to break them. I was part of the team TheRealIdex, who broke the most challenges (using mainly automated attacks). Together with the designers team who submitted the two winning candidates, we wrote an article [8] describing several attack techniques and designs used during the contest.

### **Encodings vs Side-Channel Attacks**

A classical technique to protect AES white boxes is to implement every operation in the form of look-up tables protected by random bijections called encodings. While these encodings complicate SCA, it has been proven that the latter are still effective with high probability. In an article under publication at CHES, we characterized the set of bijections that prevent existing SCA, but also showed how to break an implementation protected by these special encodings, proving that it is not possible to prevent all SCA with this technique.

## **Bibliographie**

- [1] S. Chow, P. A. Eisen, H. Johnson, and P. C. van Oorschot. A white-box DES implementation for DRM applications. ACM Workshop on Digital Rights Management 2002 : 1-15.
- [2] S. Chow, P. A. Eisen, H. Johnson, and P. C. van Oorschot. White-Box Cryptography and an AES Implementation. SAC 2002 : 250-270.
- [3] J. W. Bos, C. Hubain, W. Michiels, P. Teuwen. Differential Computation Analysis : Hiding Your White-Box Designs is Not Enough. CHES 2016 : 215-236.
- [4] P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. CRYPTO 1999 : 388-397.
- [5] D. Boneh, R. A. DeMillo, and R. J. Lipton. On the importance of checking cryptographic protocols for fault. EUROCRYPT 1997 : 37-51.
- [6] E. Dottax, C. Giraud, A. Houzelot. White-Box ECDSA : Challenges and Existing Solutions. CO-  
SADE 2021 : 184-201.
- [7] C. Giraud, A. Houzelot. Fault Attacks on a Cloud-Assisted ECDSA White-Box Based on the Residue Number System. FDTC 2023.
- [8] G. Barbu, W. Beullens, E. Dottax, C. Giraud, A. Houzelot, C. Li, M. Mahzoun, A. Ranea, J. Xie. ECDSA White-Box Implementations : Attacks and Designs from CHES 2021 Challenge. CHES 2022(4) : 527-552.

# CHASSE AUX MINES SOUS-MARINES À L'AIDE D'APPRENTISSAGE PAR RENFORCEMENT MULTI-AGENTS

**Adrien JÉGOUREL**

**Direction : Serge CHAUMETTE**  
**Encadrement : Olivier MARCEAU (Naval Group)**

*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

**L**a chasse aux mines est un enjeu majeur de la sécurité maritime. Les progrès technologiques récents ont donné lieu à l'utilisation de systèmes multi-robots autonomes, en particulier sous-marins. Cependant, les problèmes de coordination multi-agents restent un obstacle majeur à leur déploiement, notamment sous l'eau où la communication est difficile. Nous proposons d'étudier dans cette thèse des architectures basées sur des réseaux de neurones pour la prise de décision décentralisée, notamment via des algorithmes d'apprentissage par renforcement multi-agents (MARL).

## Objectifs de recherche

### Contexte

Les mines navales sont redoutables : leur menace est d'une part physique, avec des centaines de navires ou sous-marins coulés à leur actif, mais aussi psychologique, leur simple suspicion suffisant à fermer des passages maritimes entiers, ce qui peut se révéler dévastateur dans des lieux stratégiques tels que le détroit d'Ormuz.

On comprendra donc facilement l'importance qu'a pris la "guerre des mines" depuis le début du XX<sup>e</sup> siècle, notamment le volet des contre-mesures. Une approche récente, rendue nécessaire par l'apparition de mines toujours plus sophistiquées, est la chasse aux mines. Elle consiste à détecter les mines, les identifier, puis les neutraliser. La phase la plus difficile est de loin la détection, qui se base principalement sur des sonars pour trouver les objets suspects au fond de l'eau.

Les opérations de chasse aux mines ont vu grandir l'utilisation de robots sous-marins, moins coûteux et dangereux pour le personnel. Ils sont le plus souvent télé-opérés, mais la tendance est à l'autonomisation. En effet, un vif intérêt est porté par les différentes marines aux systèmes multi-robots (MRS) autonomes, plus résilients et plus efficaces.

Cependant, de manière générale les MRS font face à de nombreuses difficultés. En particulier, la coordination optimale des différents robots se ramène souvent à des problèmes d'optimisation combinatoire, comme par exemple les multiples voyageurs de commerce (mTSP), notoirement NP-difficile tout comme le TSP classique. Malheureusement, la réalité est encore plus complexes : chaque robot peut avoir différentes tâches à accomplir, qui dépendent potentiellement de la réalisation des tâches des autres robots, le tout en devant prendre les décisions de manière décentralisée, avec des com-

munications peu fiables. En effet, le moyen principal pour transmettre de l'information sous l'eau est via des ondes acoustiques, ce qui rend la perte de messages probable et limite fortement la bande passante. Une autre contrainte est la résolution *en ligne* de ces problèmes de décision afin d'être réactif face aux imprévus.

Dans une certaine mesure, il est possible de développer des algorithmes heuristiques : les systèmes actuels sont le plus souvent basés sur des parcours simples et calculés à l'avance, par exemple des trajectoires en boustrophédon. Cependant, une approche récente et prometteuse est l'utilisation de l'apprentissage par renforcement multi-agents (MARL) profond.

### Objectifs

Le problème que l'on cherche à résoudre est le suivant. On délimite une zone rectangulaire  $S$ , discrétisée en  $C_L \times C_W$  cellules. Typiquement,  $S$  est de l'ordre de plusieurs km<sup>2</sup>, et les cellules environ de  $5 \times 5$  mètres. Il y a un nombre inconnu de mines  $n_m$ , et on suppose que chaque cellule ne peut contenir au plus qu'une mine. On suppose aussi que les mines ne peuvent se trouver que sur le fond. Le relief sous-marin ainsi que le type de sol est connu à l'avance, et les robots sous-marins (AUVs) se déplacent à profondeur constante.

On dispose de  $n_d$  AUVs *détecteurs*, dont le sonar puissant mais imprécis permet de détecter les échos suspects, et  $n_{id}$  AUVs *identificateurs*, dont le sonar plus petit mais plus précis permet de déterminer si un écho suspect est une mine ou non. La fonction de détection du sonar est affectée à la fois par la distance au fond ainsi que son type (sable, rocher...). Les différents agents peuvent communiquer s'ils sont suffisamment proches (rayon de communication  $d_{com}$ ), cependant une probabilité de perte de message croissante en fonction de la distance est appliquée.

La métrique que l'on cherche à optimiser est la probabilité d'avoir détecté toutes les mines dans la zone au



bout du temps imparti  $T$ . Une illustration simplifiée de ce scénario est visible figure 17.1.

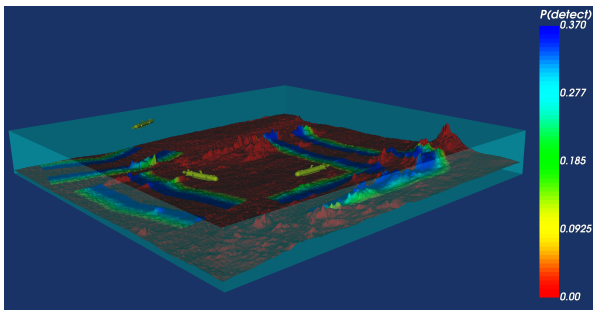


FIGURE 17.1 – Visualisation de la probabilité de détection d'une mine avec 3 AUVs, relief sous-marin au large de St-Malo

## Méthodologie

### Apprentissage par renforcement

Les algorithmes d'apprentissage par renforcement permettent d'apprendre des solutions à des problèmes de prise de décision séquentiels via des interactions répétées avec un environnement. Ce dernier est souvent modélisé par un Processus de Décision Markovien (MDP), ou une généralisation.

Pour un agent, cela signifie que l'on a un ensemble des états possibles  $\mathcal{S}$ , un ensemble des actions possibles  $\mathcal{A}$ , une fonction de transition d'un état à un autre avec une action donnée  $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ , et la fonction de récompense associée  $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ . On définit un épisode comme une séquence  $(s_t, a_t, r_t, s_{t+1}, a_{t+1}), t = 0, \dots, T, T \in \mathbb{N} \cup \{+\infty\}$ .

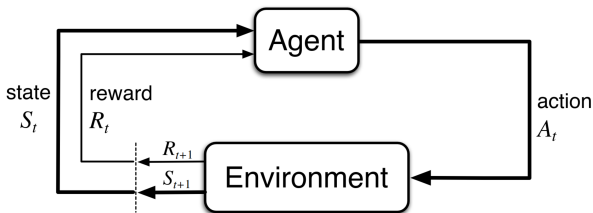


FIGURE 17.2 – Interaction agent-environnement dans un processus de décision markovien [1]

Le but de l'agent va être de trouver une *policy* optimale  $\pi : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1], \pi(s, a) = P(a|s)$  maximisant son espérance de récompense au cours d'un épisode  $\mathbb{E}_\pi \left[ \sum_{t=0}^T \gamma^t r_t \right]$ , avec  $\gamma \in [0, 1]$  un facteur d'amortissement. De nos jours,  $\pi$  est presque toujours définie à l'aide de réseaux de neurones profonds, qui permettent d'avoir des espaces d'états beaucoup plus complexes de par leur capacité d'approximation.

### Implémentation

Il existe différents algorithmes permettant d'apprendre  $\pi$ , comme PPO [2] qui est l'un des plus utilisés. Cependant, ils sont assez complexes à implémenter de manière correcte, d'autant plus si l'on veut pouvoir paralléliser les calculs, ce qui oblige en pratique à utiliser des bibliothèques dédiées. Cependant, le domaine du

MARL étant encore récent, il n'y en a pas faisant consensus, chacune proposant des fonctionnalités différentes et n'étant pas forcément bien maintenue. Une étape importante de la thèse a donc été d'en tester plusieurs, avant de retenir RLLib [3].

Ensuite, il n'existe pas dans la littérature d'approche "complète" de système multi-robots pour la chasse aux mines : beaucoup se concentrent sur un point précis (i.e. couvrir la zone) mais vont faire des hypothèses extrêmement fortes (i.e. supposer une communication parfaite, ou des agents homogènes). C'est en cela que nous pensons que le MARL peut être pertinent, en permettant de proposer une autonomie décisionnelle intégrant le plus de contraintes réalistes possible.

Cela passe premièrement par la création d'un simulateur adapté pour gérer l'environnement. Ensuite, l'ajout d'architectures novatrices, comme HetGPPO [4], qui utilise des *Graph Neural Networks* (GNN) pour représenter les interactions entre agents, ces derniers apprenant les paramètres séparément. Nous explorons également des approches avec du *curriculum learning* pour apprendre sur des scénarios de plus en plus complexes, et du *hierarchical RL* pour dissocier les stratégies de planification haut-niveau et bas-niveau.

### Résultats attendus

Nous pensons pouvoir performer sur plusieurs points par rapport aux approches actuelles. Premièrement, au niveau de la rapidité de la détection de toutes les mines d'une zone à un certain seuil de confiance. Ensuite, la capacité de résilience et d'adaptation face aux conditions difficiles : panne d'un ou plusieurs robots, très mauvaises communications. Enfin, l'adaptabilité des scénarios suivant les besoins, comme par exemple pouvoir faire remonter régulièrement les informations à un opérateur humain.

Une nuance à apporter toutefois est l'absence de garanties autres que statistiques lors de l'exécution d'une mission, puisque les décisions sont prises par un réseau de neurone qui agit comme une "boîte noire". Le respect de contraintes fortes est un des domaines de recherche du RL mais n'est pas encore mature, encore moins pour les systèmes multi-agents.

## Bibliographie

- [1] R. S. Sutton and A. G. Barto. (2018) Reinforcement learning : An introduction. MIT press.
- [2] J. Schulman et al. (2017). Proximal policy optimization algorithms. arXiv :1707.06347.
- [3] E. Liang et al. (2018) In Proceedings of the 35th International Conference on Machine Learning.
- [4] M. Bettini et al. (2023). Heterogeneous Multi-Robot Reinforcement Learning. In Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems.

# TOWARDS SOCIO-CULTURAL ARTIFICIAL AGENTS

Grgur KOVAČ

Direction : Pierre-Yves Oudeyer  
Co-Direction : Peter Ford Dominey

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

We explore the interplay between human culture and artificial intelligence. More precisely, we study how can research conducted in the field of psychology be useful in AI. We explore this question in two projects. In the *SocialAI School*, we follow developmental theories from Jerome Bruner and Michael Tomasello to outline the most basic core concepts and socio-cognitive abilities relevant to study how a system enters a culture. Then, we present a tool facilitating research in this domain as well as a series of case studies. In *Stick to Your Role!*, we explore the cultural expression in Large Language Models (LLMs). We leverage the psychological methodology regarding intrapersonal (Ipsative) and interpersonal (Rank-order) value stability. We study how context influences the expression of basic personal values in different LLMs and observe similar trends in terms of value stability expressed by different models.

## The SocialAI School

### Introduction

Developmental psychologists have long-established socio-cognitive abilities as a core aspect of human intelligence and development [2]. Those abilities enable us to enter, participate in and benefit from human culture. Humans are then able to push our culture forward by improving on already existing cultural artifacts. This is referred to as the cumulative cultural evolution, and it has been argued that the most impressive of human achievements are the product of it [3]. It seems clear that to construct artificial agents capable of interacting with and learning from humans one must equip them with socio-cognitive abilities enabling them to enter the human culture. This would enable artificial agents to benefit from our culture, and also push it forwards, i.e. to participate in the cumulative cultural evolution.

### The SocialAI School

In this project, we follow the work of Michael Tomasello and Jerome Bruner who outlined a set of core socio-cognitive skills, e.g. social cognition (joint attention, theory of mind), referential and conventionalized communication, imitation, role-reversal, scaffolding, and many more [4, 2]. Importantly, they also showed the importance of the (cultural) environment for cognitive development.

To introduce some of those concepts to the AI community, we created a tool for procedural generation of environments - The SocialAI school. With The SocialAI school, experiments studying those socio-cognitive abilities can be easily conducted and cognitive-science experiments reconstructed. Furthermore, The SocialAI School enables us to generate both multimodal environments (suited for RL agents) and pure text version of those environments (suited for LLM-based agents). An

example of a SocialAI school environment is shown in figure 18.1. In it, the peer is pointing towards the red box. The agent (the red triangle) has to infer this to mean that the apple is hidden inside the red box.

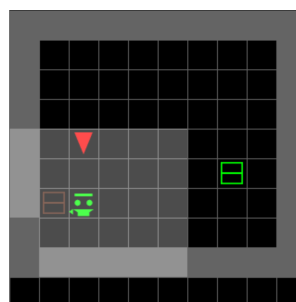


FIGURE 18.1 – An example of an environment that can be created with The SocialAI school. The green peer is pointing toward the brown box. The task of the agent (the red triangle) is to infer this to mean that the brown box contains an apple.

### Experiments

We conducted many experiments, here we outline a few more important ones. We experimented with multi-modal RL agents. We tested generalization of inferring the meaning of referential communication (ex. the pointing gesture) to new scenarios/objects. We found that such a generalization is very hard for standard reinforcement learning agents. We show how a scaffolded environment helps with learning complex interaction sequences (formats). To show how cognitive science experiments can be recreated - we reconstruct a study of role reversal from [5]. Furthermore, we conducted experiments regarding other aspects of social-cognition : joint attention, imitation, perspective taking, etc. We experimented with LLM-based interactive agents. We show that a simple LLM-based agent can achieve some performance, but still fails to generalize. This moti-

vates future work for creating more complex LLM-based agents.

## Stick to your Role! Stability of Personal Values Expressed in Large Language Models

### Introduction

There has been a growing body of research using Large Language Models (LLMs) to simulate individuals or human populations. Those studies usually focus on how a model can express some behavior or values and often overlook the underlying problem that LLMs are highly context-dependent. In this project, we study the robustness of LLMs to seemingly unrelated perturbations in the context. Instead of evaluating models on some behavior (and testing robustness along the way), we study how a model's behavior changes over contexts as a primary question. We argue that context-dependence, and more precisely, value stability, should be seen as another dimension of LLM comparison alongside others such as knowledge, model size, speed. In other words, we study and compare the value stability of LLMs along different contexts.

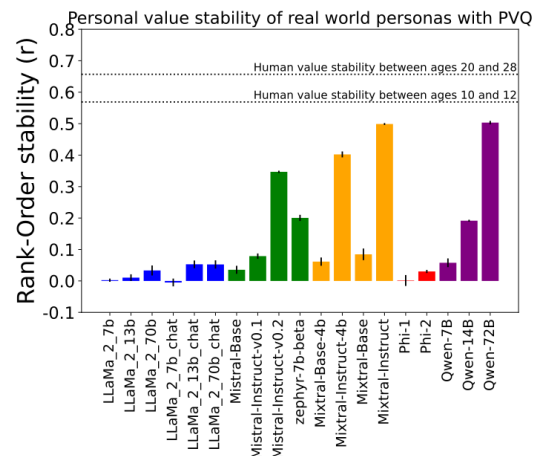
### Methods

We leverage the PVQ questionnaire [6] associated with the Schwartz Theory of Basic Values [7], which defines ten values : Self-Direction, Stimulation, Hedonism, Achievement, Power, Security, Conformity, Tradition, Benevolence, Universalism. We study the stability of values expressed by LLMs without defining a specific persona to simulate, and with instructing the model to simulate various well-known fictional and real-world individuals (when simulating populations). Different contexts are induced by simulating conversations on different topics : correcting grammar, writing a poem, playing chess, answering a history question and inventing a joke. Furthermore, we conduct experiments on three downstream tasks of varying difficulty. Following psychological methodology, we evaluate two types of value stability : Rank-order, Ipsative. Rank-order stability refers to the correlation of participant ranks in two different context, i.e. interpersonal stability. Ipsative stability refers to the correlation of the order of values in the same individual in different context - intrapersonal stability.

### Experiments

We observed consistent trends of value stability : Mixtral, Mistral and Qwen model families were more stable (these trends are also observed on downstream behavioral tasks). LLMs instructed to simulate personas exhibit much lower than human stability (despite the comparison being skewed in their favor), which further diminishes over longer conversations. This insight motivates future research on models specialized in simulating coherent populations of individuals. More precisely, these results further push the point that benchmarks and

psychological questionnaires administered to LLMs cannot be used and interpreted in the same way as with humans, i.e. much more attention must be given to studying context-dependence. Most importantly, they imply that rather than evaluating many questions from a single context (as is currently common), one should also evaluate the same questions from many different contexts. To our knowledge, we propose the first analysis of value stability in LLMs.



**FIGURE 18.2** – Rank-order stability of different Large Language Models. Mistral, Mixtral and Qwen families exhibit higher stability. However this stability is still much lower than that observed during human development (despite the comparison being skewed in favor of LLMs)

## Bibliographie

- [1] Vygotsky, L. & Cole, M. Mind in society : the development of higher psychological processes. ( Harvard University Press Cambridge ,1978)
- [2] Bruner, J. Child's Talk : Learning to Use Language. Child Language Teaching And Therapy. **1**, 111-114 (1985), [https ://doi.org/10.1177/026565908500100113](https://doi.org/10.1177/026565908500100113)
- [3] Tomasello, M. The Cultural Origins of Human Cognition. (Harvard University Press,1999), <http ://www.jstor.org/stable/j.ctvjjsf4jc>
- [4] Tomasello, M. Becoming human. Becoming Human. (2019)
- [5] Fletcher, G., Warneken, F. & Tomasello, M. Differences in cognitive processes underlying the collaborative activities of children and chimpanzees. Cognitive Development. **27**, 136-153 (2012), <https ://www.sciencedirect.com/science/article/pii/S08852014>
- [6] Cieciuch, J. & Schwartz, S. The Number of Distinct Basic Values and Their Structure Assessed by PVQ-40. J. Pers. Assess. **94**, 321-328 (2012,5)
- [7] Schwartz, S. An overview of the Schwartz theory of basic values. Online Readings In Psychology And Culture. **2**, 11 (2012)

# MÉTHODES DE RÉDUCTION D'ORDRE-PARTIEL POUR LA VÉRIFICATION DE SYSTÈMES CONCURRENTS TEMPS-RÉEL

Sarah LARROZE

Direction : Frédéric HERBRETEAU  
Co-Direction : Igor WALUKIEWICZ

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Un système concurrent est composé de différents processus qui s'exécutent en parallèle et peuvent se synchroniser via des actions communes. On peut ajouter des contraintes temporelles sur ce type de système et ainsi obtenir un système concurrent temps-réel. Un système concurrent temps-réel peut être représenté par un réseau d'automates temporisés et l'exécution du système correspond au produit de ce réseau d'automates temporisés. On peut réduire la vérification de systèmes concurrents temps-réel à un problème d'accessibilité d'un état final dans un automate temporisé. Dans la pratique, les systèmes considérés sont très grands et il est trop coûteux de calculer cet automate produit car le nombre d'états accessibles peut être exponentiel en la taille du système. Cette explosion d'états constitue le principal obstacle à la résolution de problèmes d'accessibilité dans les systèmes concurrents. Par exemple, si on considère un système constitué de  $n$  processus qui exécutent chacun une action sans se synchroniser, il y a  $n!$  exécutions possibles dans ce système, correspondant à tous les entrelacements possibles des actions des différents processus. Cependant, comme les différents processus ne se synchronisent pas, l'exécution d'un processus donné n'affecte pas les autres processus et alors les  $n!$  exécutions du système aboutissent toutes dans un même état final. Ainsi, toutes ces exécutions peuvent être considérées comme équivalentes et il serait alors suffisant de n'en explorer qu'une seule pour répondre au problème d'accessibilité d'un état final. En se basant sur cette idée, des méthodes de réduction d'ordre-partiel ont été introduites dans les années 90 [2, 9, 1]. Ces méthodes définissent une notion d'indépendance entre les actions du système et ont pour but de réduire la portion du système qui est explorée en évitant d'explorer des exécutions qui sont équivalentes. Ces méthodes se sont révélées très efficaces sur certains systèmes et l'objectif de cette thèse est de les mettre en œuvre sur des systèmes concurrents temps-réel.

## Méthodes de réduction d'ordre-partiel

Un système concurrent peut être représenté par un réseau d'automates finis dans lequel chaque processus est représenté par un automate fini. Un processus peut effectuer deux types d'actions : des actions locales qui s'effectuent localement à ce processus et des actions de synchronisation qui impliquent au moins deux processus. Un état du système correspond à un tuple contenant une localisation pour chaque processus et une transition est exécutable dans un état donné si elle est exécutable dans la localisation de chaque processus impliqué dans cette transition, quand le système est dans cet état. Un état final du système consiste en un état qui ne contient que des localisations finales des processus du système. Soit  $a$  une action, on dénote par  $dom(a)$  l'ensemble des processus qui utilisent cette action. La notion d'indépendance que nous avons définie sur ce type de systèmes est la suivante : deux actions  $a$  et  $b$  sont indépendantes si :

$$dom(a) \cap dom(b) = \emptyset$$

Les traces d'exécution du système correspondent à des séquences d'actions. Étant données deux traces d'exécution  $\rho$  et  $\rho'$ , ces traces sont dites équivalentes, noté

$\rho \sim \rho'$ , si on peut obtenir  $\rho'$  à partir de  $\rho$  en permutant des actions indépendantes adjacentes. On peut alors partitionner l'ensemble des exécutions du système en classes d'équivalence et il suffit d'explorer une seule trace par classe d'équivalence pour répondre au problème d'accessibilité d'un état final.

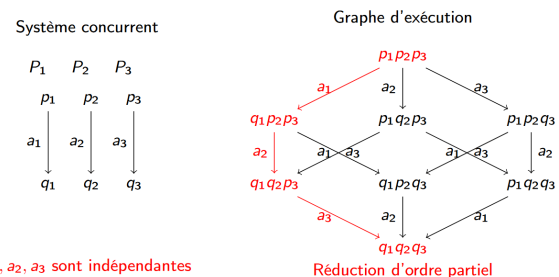


FIGURE 19.1 – Un système concurrent dans lequel toutes les actions sont indépendantes. On peut voir sur le graphe d'exécution de ce système que toutes les exécutions sont équivalentes, il suffit alors de n'en explorer qu'une pour répondre au problème d'accessibilité d'un état final.

Si exactement une trace par classe d'équivalence est explorée, la méthode de réduction d'ordre-partiel est dite optimale [4, 5, 6].



## Systèmes concurrents temps-réel

Un système concurrent temps-réel est un système concurrent dans lequel les processus possèdent des contraintes temporelles qui restreignent leur exécution. Ce type de systèmes peut être modélisé par un réseau d'automates temporisés dans lequel chaque processus est représenté par un automate temporisé. Un automate temporisé est un automate fini dans lequel des variables réelles nommées horloges sont ajoutées. Ces horloges évoluent en fonction du temps et leur valeur peut être testée par une garde ou remise à zéro. La sémantique usuelle pour les automates temporisés [3] considère un temps global pour tous les processus, c'est-à-dire que toutes les horloges évoluent selon le même temps global dans tous les processus. La Figure 19.2 présente un système concurrent temps-réel. Ici,  $x$  est une horloge du processus  $P_1$  et  $y$  est une horloge du processus  $P_2$ . Dans l'état initial, les deux horloges valent 0. Dans l'exécution en temps global, on observe qu'après l'exécution de  $a$  depuis l'état initial, l'action  $b$  ne peut pas être exécutée. En effet, on a fait passer un délai d'au moins 2 secondes pour exécuter  $a$  et alors les horloges  $x$  et  $y$  valent plus que 2 donc la garde  $y < 1$  n'est pas satisfaite et  $b$  ne peut pas être exécutée. Il est cependant possible d'exécuter d'abord l'action  $b$  puis l'action  $a$ . Ainsi, l'indépendance entre les actions  $a$  et  $b$  est perdue à cause des contraintes temporelles, bien que  $a$  et  $b$  soient des actions locales à chaque processus. On ne peut donc pas appliquer de méthodes de réduction d'ordre-partiel dans la sémantique en temps global.

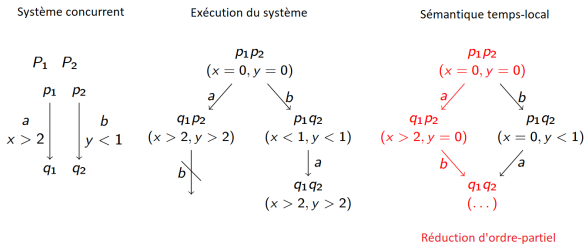


FIGURE 19.2 – Un système concurrent temps-réel avec son exécution dans la sémantique en temps global et dans la sémantique en temps local.

Récemment, une sémantique en temps local a été introduite [7]. Dans cette sémantique, on considère un temps local à chaque processus, c'est-à-dire qu'il y a un temps qui s'écoule indépendamment dans chaque processus. Ce temps est régi par une horloge de référence et chaque processus possède sa propre horloge de référence. Afin d'effectuer une action commune, les horloges de référence des processus impliqués dans cette action doivent se synchroniser. Sur la Figure 19.2, on peut observer qu'on récupère l'indépendance entre les actions  $a$  et  $b$  sur l'exécution du système dans la sémantique en temps local. En effet, quand on exécute l'action  $a$ , le temps ne s'est écoulé que dans le processus  $P_1$  et la garde  $y < 1$  est alors satisfaite après l'exécution de  $a$  puisque le temps ne s'est pas écoulé dans le processus  $P_2$ . Ainsi, on va pouvoir appliquer des méthodes

de réduction d'ordre-partiel quand on travaille avec la sémantique en temps-local.

## Méthodologie

Dans le cadre de cette thèse, nous avons défini des méthodes de réduction d'ordre-partiel statiques, c'est-à-dire que les actions à exécuter dans chaque état du système sont fixées avant l'exécution du système [2, 1, 9]. Nous avons également travaillé sur des méthodes de réduction d'ordre-partiel dynamiques [8, 5, 4], c'est-à-dire qu'on décide pendant l'exécution du système quelles actions doivent être exécutées dans chaque état, en se basant sur ce qui a déjà été exploré. Ces méthodes de réduction d'ordre-partiel ont été implémentées dans l'outil TChecker et ont été testées sur des benchmarks de la littérature.

## Bibliographie

- [1] Edmund M. Clarke and Orna Grumberg and Marius Minea and Doron A. Peled. State Space Reduction Using Partial Order Techniques. 1999.
- [2] Patrice Godefroid. Partial-Order Methods for the Verification of Concurrent Systems - An Approach to the State-Explosion Problem. 1996.
- [3] Rajeev Alur and David L. Dill. A Theory of Timed Automata. Theor. Comput. Sci., vol. 126, p183–235. 1994.
- [4] Michalis Kokologiannakis, Iason Marmanis, Vladimir Gladstein, Viktor Vafeiadis. Truly stateless, optimal dynamic partial order reduction. Proc. ACM Program. Lang. vol. 6, p1–28. POPL. 2022.
- [5] Abdulla, Parosh Aziz and Aronis, Stavros and Jonsen, Bengt and Sagonas, Konstantinos. Source Sets : A Foundation for Optimal Dynamic Partial Order Reduction. ACM, p25 :1–25 :49. 2017.
- [6] Vineet Kahlon, Chao Wang, Aarti Gupta. Monotonic Partial Order Reduction : An Optimal Symbolic Partial Order Reduction Technique. CAV '09.
- [7] Govind, R. and Herbretau, Frédéric and Srivathsan, B. and Walukiewicz, Igor. Abstractions for the local-time semantics of timed automata : a foundation for partial-order methods. LICS 2022.
- [8] Cormac Flanagan and Patrice Godefroid. Dynamic Partial-Order Reduction for Model Checking Software. POPL. 2005.
- [9] Antti Valmari. A Stubborn Attack On State Explosion Computer Aided Verification, 2nd International Workshop, CAV '90.

# LES PRATIQUES DE CODE : DE LA DOCUMENTATION À LA DÉTECTION

Corentin LATAPPY

Direction : Jean-Rémy FALLERI  
Encadrement : Thomas DEGUEULE

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

**L**es pratiques de code sont de plus en plus utilisées dans le domaine du développement logiciel. Leur mise en place permet d'assurer la maintenabilité, la lisibilité et la consistance du code, ce qui contribue fortement à la qualité logicielle. La majorité de ces pratiques est implémentée dans des outils qui permettent d'alerter automatiquement les développeurs lorsqu'une pratique n'est pas respectée.

De plus en plus d'organisations créent leurs propres pratiques internes et rencontrent des problèmes sur leur adoption par les développeurs. Afin d'être appliquée, une pratique doit d'abord être comprise par les développeurs, impliquant donc d'avoir une documentation correctement rédigée. Ensuite, pour favoriser leur adoption, il faudrait pouvoir étendre les outils d'analyse existants pour y intégrer de nouvelles pratiques, ce qui est difficile compte tenu de l'expertise nécessaire pour apporter ces modifications.

## Contexte

La qualité logicielle fait maintenant partie intégrante du processus de création d'une solution logicielle, en permettant d'assurer que les produits logiciels répondent aux attentes des utilisateurs, tout en étant fiables et maintenables. Pour atteindre ces standards élevés de qualité, il est impératif de s'attarder sur ce qui constitue avant tout un logiciel : le code. En effet, la manière dont le code est conçu, écrit, testé, et maintenu joue un rôle crucial dans la réalisation des objectifs de qualité. C'est le résultat de ce mélange, alliant connaissances théoriques, expériences pratiques et évolution constante des technologies, qui a mené à l'émergence des pratiques de code.

Les pratiques de code prennent différentes formes et peuvent porter sur divers aspects inhérents à l'écriture du code. Cela peut être simplement des conventions visant à maintenir la cohérence du style du code à travers un projet, tout comme la gestion des dépendances externes pour éviter les conflits, ou bien encore la mise en place de tests garantissant que le code fonctionne comme attendu. Elles influent directement sur les aspects principaux de la qualité logicielle, telles que la maintenabilité, la performance et la sécurité du logiciel. Adopter et maintenir de bonnes pratiques de code est donc un investissement stratégique pour la qualité globale du logiciel.

Packmind<sup>8</sup>, entreprise dans laquelle j'effectue ma thèse Cifre, développe un outil appelé "Ateliers Craft". Le but de ces ateliers est de créer un temps d'échange entre les développeurs d'une entreprise afin de mettre en commun leurs pratiques de code. Cela permet de faire émerger une documentation vivante et fortement liée au code interne sur la façon de développer. L'intention

principale de Packmind est donc de diffuser et d'homogénéiser les pratiques internes afin d'améliorer de manière globale la qualité logicielle au sein d'un projet.

## Problématiques et Objectifs

Avec le recul et les différentes expériences des utilisateurs, notre solution comporte un certain nombre de problématiques liées à un manque d'opérationnalisation.

**Améliorer la qualité de la documentation des pratiques.** La problématique principale concerne la bonne transmission de la connaissance à un développeur qui découvre une nouvelle pratique. Avec Packmind, cette connaissance est transmise en théorie pendant le déroulement des Ateliers Craft. Cependant, si on considère le cas de nouveaux arrivants dans une équipe, n'ayant donc pas pu participer aux précédents ateliers, ils n'auront pas d'autre choix que d'aller consulter la documentation des pratiques déjà créées s'ils veulent être informés de manière autonome. On rencontre le même cas lorsqu'un développeur utilisant un linter dans son IDE est alerté qu'il a écrit du code qui ne respecte pas une pratique. Lorsqu'il est notifié d'une telle violation, le développeur voudra corriger son erreur pour obtenir un code conforme. S'il a déjà rencontré cette règle, cette opération ne lui posera aucun problème, mais si ce n'est pas le cas, il va là aussi parcourir la documentation de la pratique associée. Cette documentation devient donc un point d'entrée capital pour apporter la connaissance aux développeurs, ainsi que pour promouvoir le respect de ces pratiques. Dans cette thèse nous souhaitons donc fournir des recommandations à destination des rédacteurs de documentation. Nous souhaitons d'abord établir un état des lieux sur la façon dont sont docu-

8. <https://www.packmind.com>



mentées les pratiques de code, puis ensuite évaluer les attentes des développeurs vis-à-vis de ce type de documentation.

**Apprendre automatiquement les pratiques internes.** La seconde problématique est le suivi de la mise en application des pratiques par les développeurs. En effet, les Ateliers Craft sont des outils idéaux pour faire émerger les pratiques, mais pas pour les faire appliquer à l'échelle d'un projet. Certes les pratiques sont validées en accord pendant les ateliers, mais leur nombre grandissant fait qu'il devient difficile de toutes s'en souvenir. De plus, le turnover important dans le domaine de l'ingénierie logicielle implique un changement régulier de développeurs au sein d'une équipe à qui il faut inculquer ces pratiques. À l'image d'un linter, nous souhaitons donc fournir un outil qui alertera un développeur lorsqu'une pratique ne sera pas respectée. Il ne devra nécessiter aucune intervention humaine lorsqu'une nouvelle pratique est ajoutée à la base de connaissances, et idéalement être indépendant du langage de programmation.

## Contributions

Pour répondre à la problématique sur la rédaction d'une pratique de code, nous avons souhaité étudier des outils existants et déjà utilisés par les développeurs, à savoir les linters. Ces outils sont très largement utilisés par les développeurs pour diffuser et faire appliquer les pratiques de code. Ils représentent un atout majeur lorsqu'ils sont directement intégrés dans les IDE puisque la détection a lieu en simultanée avec l'écriture du code. Lorsqu'une violation est détectée, le linter alerte le développeur à travers une notification. Cette notification affiche le minimum d'informations et il est donc nécessaire de consulter la version en ligne afin d'obtenir plus de détails. Nous pouvons donc utiliser ces exemples de documentation afin de réaliser un état des lieux sur leur composition. Nous avons ainsi analysé la documentation de plus de 100 règles de 16 linters pour 7 langages, et nous en avons extrait une taxonomie sur les objectifs (Quoi, Pourquoi, Correction) et les types de contenu (Texte, Code, Lien) dans une documentation. Nous avons ensuite confronté ces résultats en menant une enquête auprès des développeurs. Cela nous a permis de mettre en avant des problèmes avec la documentation du Pourquoi, considéré essentiel par les développeurs, mais majoritairement absent ou de mauvaise qualité en pratique. Nous avons aussi trouvé que le Code et le Texte sont intéressants pour documenter le Quoi et la Correction. Ce travail a contribué, en plus de son package de réplique [2], à la publication suivante :

C. Latappy, T. Degueule, J.-R. Falleri, R. Robbes, X. Blanc, C. Teyton, "What the Fix? A Study of ASATs Rule Documentation", 46th International Conference on Program Comprehension (ICPC), Lisbon, Portugal, 2024, doi : 10.1145/3643916.3644404

Nous avons également créé un outil web<sup>9</sup> permettant d'accompagner notre publication en fournissant plus de données, et notamment de consulter les résultats de l'enquête de manière interactive.

Concernant la seconde problématique, nous avons voulu répondre à la question de savoir si les pratiques de code pouvaient être apprises automatiquement à partir d'exemples de code. Nous avons ainsi réalisé une étude de faisabilité à l'aide de CodeBERT, modèle de ML, pour tenter d'apprendre des pratiques et de les identifier par la suite. La spécificité de ce travail a été de se placer dans le contexte que nous avons chez Packmind. Ce contexte, inhabituel pour le machine learning, nous a notamment permis d'étudier l'impact du nombre d'exemples sur la qualité de l'apprentissage. Pour procéder à cet apprentissage, nous avons généré un dataset [1] contenant quasiment 13 millions de violations pour 38 règles différentes du linter ESLint. Cela nous a permis de réaliser diverses expériences mixant les tailles d'apprentissage et les ratios entre du code conforme, non-conforme et aléatoire. Bien que les résultats dans le cadre expérimental soient prometteurs, nous obtenons des résultats de précision trop faible pour que notre solution soit utilisable dans un cas industriel. Cette étude a mené à cette seconde publication :

C. Latappy, Q. Perez, T. Degueule, J.-R. Falleri, C. Urtado, S. Vauttier, X. Blanc, C. Teyton, "MLinter : Learning Coding Practices from Examples—Dream or Reality?", 2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), Taipa, Macao, 2023, doi : 10.1109/SANER56733.2023.00092

## Bibliographie

- [1] C. Latappy et al., « Dataset - MLinter : Learning Coding Practices from Examples—Dream or Reality ? » Zenodo, 20 novembre 2022. doi : 10.5281/zenodo.7341456.
- [2] C. Latappy, T. Degueule, J.-R. Falleri, R. Robbes, X. Blanc, et C. Teyton, « Replication Kit - What the Fix? A Study of ASATs Rule Documentation ». Zenodo, 17 janvier 2024. doi : 10.5281/zenodo.10522473.

9. <https://icpc2024-asats.github.io>

# RECONFIGURATION AND KEMPE CHANGES

Clément LEGRAND-DUCHESNE

Direction : Marthe BONMAY

Co-Direction : Vincent DELECROIX

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

This thesis focuses on combinatorial reconfiguration and structural problems with a geometric flavor. Reconfiguration is the study of configuration spaces with respect to an elementary operation transforming a configuration into another. Famous reconfiguration problems include puzzle games such as Rubik's Cube or Sokoban for example, or problems on graphs such as reconfiguration of colorings. In this thesis, we study several reconfiguration operations, such as Reidemeister moves of knot diagrams or a generalisation of the Rubik's Cube to square-tiled surfaces. In this extended abstract, we focus on another operation : Kempe changes on colorings.

## Kempe changes

In 1852, Guthrie observed while coloring maps of counties of England that it was possible to color them using no more than four colors, ensuring that no adjacent counties shared the same color. He wondered if this phenomenon could be extended to all planar maps, giving rise to the renowned Four-Color-Problem : are all planar graphs 4-colorable (a  $k$ -coloring of a graph  $G$  is a map that assigns to each vertex of  $G$  a color in  $\{1, \dots, k\}$ , such that adjacent vertices have distinct colors) ?

In 1879, Alfred Kempe<sup>10</sup> introduced an elementary modification of a graph coloring as a tool toward solving this problem. To change the color of a vertex  $u$  from  $a$  to  $b$  and obtain a proper coloring, the neighbors of  $u$  colored  $b$  need to be recolored. By recoloring them with  $a$  and propagating the swapping of  $a$  and  $b$  recursively, we obtain another proper coloring, with  $u$  colored  $b$ . Formally, given a  $k$ -colored graph  $G$  and two colours  $a$  and  $b$ , an  $\{a, b\}$ -Kempe chain is a connected component in the subgraph induced in  $G$  by the vertices colored  $a$  or  $b$ , and Kempe change consists in swapping the two colors within a Kempe chain, thereby obtaining a new  $k$ -coloring of the graph (see Figure 21.1).

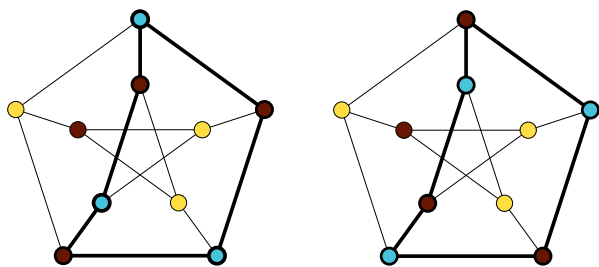


FIGURE 21.1 – Two 3-colorings of the Petersen graph that differ from one Kempe change. The corresponding  $\{\bullet, \bullet\}$ -Kempe chain is thickened.

Using this notion, we can easily prove that all planar graphs are 5-colorable and while Kempe's proof of

the Four-Color-Theorem turned out to be false, Kempe changes are a crucial tool in the computer assisted proof of Apple and Haken. Although this proof was since shortened by Robertson, Sanders, Seymour and Thomas, it still uses Kempe changes and computer-check. Finding a shorter proof of the Four-Color-Theorem is still an active field of research.

When considering edge-colorings instead of vertex-colorings, Kempe chains are a lot more structured as they are either paths or cycles. The chromatic index  $\chi'$ , which is the number of colors needed to color the edges of a graph, is trivially lower bounded by the maximum degree  $\Delta$  and Vizing proved that all graphs are  $(\Delta + 1)$ -edge-colorable. Kempe changes are here again decisive in his proof as no proof avoiding this tool is known as of today.

Two  $k$ -colorings of a graph are Kempe equivalent if one can be obtained from the other through a series of Kempe changes. Vizing proved that all the edge-colorings of a graph of maximum degree  $\Delta$  are equivalent to a  $(\Delta + 1)$ -edge-coloring up to a sequence of Kempe changes. Vizing conjectured that in fact, every  $k$ -edge-coloring of  $G$  should be Kempe equivalent to an optimal coloring, that is a  $\chi'(G)$ -edge-coloring. In particular, this conjecture implies that all  $(\chi'(G) + 1)$ -edge-colorings of  $G$  are Kempe equivalent. Recently, this conjecture was confirmed, first by Bonamy, Defrain, Klimošová, Lagoutte and Narboni on triangle free graphs and then by Narboni on all graphs.

## Main questions

The same questions can be asked on vertex-colorings : under which conditions is a graph  $k$ -recolorable, that is all its  $k$ -colorings are Kempe equivalent ? Is there an optimal coloring in all the Kempe equivalence classes ; is there an efficient algorithm to find it ? Our main interest is the study of the reconfiguration graph  $\mathcal{K}_k(G)$ , whose vertices are the  $k$ -colorings of  $G$  and in which

10. see [1, 2] and references therein for the complete bibliography

two colorings are adjacent if and only if they differ by one Kempe change. The former questions can be restated as follows : in which setting is the reconfiguration graph connected, that is, any two  $k$ -colorings are Kempe equivalent? When this is the case, can we bound the length of the shortest sequence of Kempe changes between any two colorings, i.e. the diameter of the reconfiguration graph? Beyond the intrinsic interest of theoretical results, Kempe changes are motivated by practical applications in statistical physics, and closer to graph theory, Kempe equivalence can be studied with a goal of obtaining a random coloring. Random sampling is also deeply related to approximate counting the colorings of a graph. Efficient random sampling can sometimes be achieved by applying random walks and rapidly mixing Markov chains. In particular, the Wang-Swendsen-Kotecký dynamics performs random Kempe changes at each step and is studied in statistical mechanics to understand the transition phase of the Potts model.

## Recoloring sparse graphs

### Planar graphs

In 1975, Meyniel proved that the  $k$ -colorings of any planar graph are Kempe equivalent for  $k \geq 5$ . This is tight as there exists planar graphs with frozen 4-colorings : colorings in which any two colors induce a connected subgraph, hence on which Kempe changes do not modify the partition induced by the color classes (see Figure 21.2). However, his proof produces a reconfiguration sequence of length exponential in the number of vertices. With Quentin Deschamps, Carl Feghali, František Kardoš and Théo Pierron[2], we strengthened this result by proving that for any planar graph  $G$  and any  $k \geq 5$ , the Kempe reconfiguration graph  $\mathcal{K}_k(G)$  has a polynomial diameter.

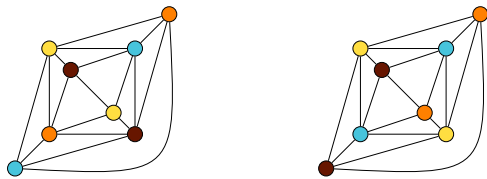


FIGURE 21.2 – Two frozen 4-colorings of a planar graph.

### Degenerate graphs

In 1978, Las Vergnas and Meyniel studied the Kempe equivalence in  $d$ -degenerate graphs. They proved that all  $d$ -degenerate graphs are  $k$ -recolorable for  $k \geq d + 1$ . Once again, their proof gives an exponential bound on the number of Kempe changes needed to go from one coloring to another, and finding a polynomial bound on the diameter of  $\mathcal{K}_k(G)$  in this setting is still a widely open problem.

### Graphs of bounded degree

Any graph  $G$  can be greedily colored with  $(\Delta + 1)$  colors, where  $\Delta$  is the maximum degree of  $G$ . Brooks' theorem

states that if  $G$  is not a clique or an odd cycle, then  $\Delta$  colors suffice. Many different proofs of this theorem exist, some of them using Kempe changes.

Mohar conjectured that all the  $k$ -colorings of a graph are Kempe equivalent for  $k \geq \Delta$ . Note that the result of Las Vergnas and Meyniel settles the case of non-regular graphs. Feghali, Johnson and Paulusma proved that the conjecture holds for all cubic graphs but the 3-prism (see Fig 21.3). Indeed, the 3-prism  $K_2 \square K_3$  admits frozen 3-colorings and hence is not 3-recolorable.

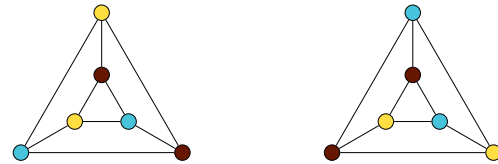


FIGURE 21.3 – Two frozen 3-colorings of the 3-prism  $K_2 \square K_3$ .

Bonamy, Bousquet, Feghali and Johnson later showed that the conjecture also holds for  $\Delta$ -regular graphs with  $\Delta \geq 4$ . Both papers heavily rely on a lemma of Las Vergnas and Meyniel and provide sequences that possibly have exponential length.

With Marthe Bonamy and Vincent Delecroix, we proved in [1] polynomial bounds on the diameter of the Kempe reconfiguration graph in three sparse graph classes. In particular, we prove that for all graph  $G$  but the 3-prism, all the  $k$ -colorings of  $G$  are Kempe equivalent up to  $O(\Delta n^2)$  Kempe changes for  $k \geq \Delta$ .

## Bibliographie

- [1] Marthe Bonamy, Vincent Delecroix, and Clément Legrand-Duchesne. Kempe changes in degenerate graphs. *European Journal of Combinatorics*, page 103802, November 2023.
- [2] Quentin Deschamps, Carl Feghali, František Kardoš, Clément Legrand-Duchesne, and Théo Pierron. Strengthening a Theorem of Meyniel. *SIAM Journal on Discrete Mathematics*, 37(2) :604–611, June 2023.

# ADVANCING LUNG IMAGING : DEEP LEARNING FOR MR-TO-CT SYNTHESIS

Arthur LONGUEFOSSE

Direction : Pascal DESBARATS

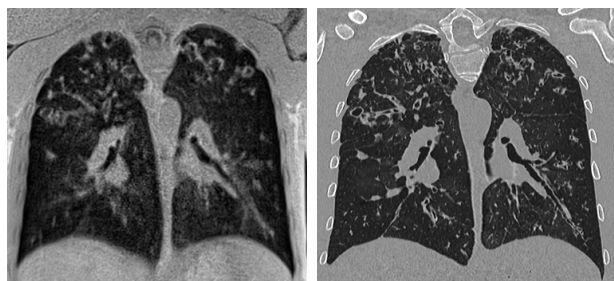
Co-Direction : Fabien BALDACCI

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

In the realm of medical imaging, the translation from MR to CT images represents a pivotal advancement, aimed at combining the non-invasive nature of MRI with the superior spatial resolution of CT, without the associated risks of ionizing radiation. This research delves into various deep learning approaches for MR-to-CT lung synthesis, exploring generative adversarial networks, perceptual loss, and structure-prioritized loss functions. Additionally, it assesses the performance of these approaches across different image representations, such as 2D, 2.5D, and 3D, and evaluates these methods against both local and global metrics, as well as their clinical relevance. The outcomes suggest promising avenues for enhancing image synthesis fidelity, emphasizing the significance of anatomical structure reconstruction.

## Introduction

Lung imaging using MRI is still a major challenge in medical imaging, owing to the low proton content, the rapid decay of lung signal due to susceptibility artifacts and respiratory motions. Still, lung MRI would be a desirable approach to computed tomography (CT) given its lack of radiation exposure and the possibility to combine both morphological and functional information. The introduction of ultrashort or zero echo-time (UTE/ZTE) MRI techniques has shown promise in capturing detailed lung images but has not seen widespread clinical adoption. This is partly because the images from UTE MRI can look different from CT images, with variations in texture and potential for blurring and noise. Moreover, the adoption of UTE MRI is limited by the need for specific MRI training for radiologists and the lack of standardized procedures. Generating CT images from MRI, particularly using UTE MRI, could address these limitations in lung imaging, offering a promising solution.



UTE MRI

Corresponding CT

**FIGURE 22.1** – Visual comparison of thoracic UTE MR and CT modalities of the same patient at a corresponding coronal slice. The CT scan shows a higher signal quality, greater contrast and sharpness, and fewer artifacts compared to the MRI.

## Methodology

In our study, we explore deep learning methods, notably generative adversarial networks (GANs), renowned for their ability to generate synthetic images closely resembling original datasets. GANs consist of two parts : a generator, which creates images, and a discriminator, which evaluates them. By learning complex patterns between MR and CT images, GANs can produce high-quality synthetic data that mimics CT images from MR inputs. Our initial studies aimed to demonstrate the feasibility of using GANs for MR-to-CT lung image synthesis. We employed the pix2pixHD architecture along with SPADE normalization to prevent the generation of false positives with a context-aware image translation. Collaborating with radiologists, we confirmed the clinical relevance of these methods and showcased that synthetic CT images generated from MRI not only exhibit almost perfect concordance with real CT images for depicting CF-related lung alterations, but also surpass the image quality of traditional UTE MRI scans [2].

### Data representation

One of our initial efforts focused on optimizing data utilization, as highlighted in our work [1]. In medical imaging, where data availability can often be a bottleneck, leveraging the full potential of existing datasets is paramount. We meticulously navigated through the essential stages of MR-to-CT image translation using GANs, covering data selection, registration, and normalization for effective learning and reconstruction.

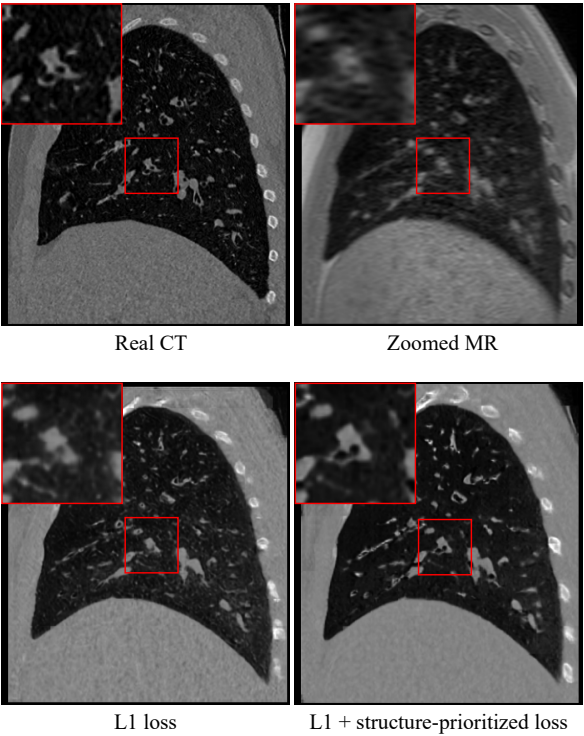
Expanding upon these findings, our most recent study introduced a structure-prioritized loss function, designed to enhance the detail and accuracy of specific anatomical features in synthesized images [5]. This approach, using feature maps from a pre-trained model on a specific segmentation task, demonstrated superior performance in reconstructing detailed structures like



airways, beyond what traditional global reconstruction methods achieve. Our findings emphasize the potential of combining these innovative loss function to refine MR to CT image synthesis, aiming for both global image quality and precise structural detail. Future directions will explore the extension of these methods to other anatomical parts and further refine the synthesis process, ensuring the broader applicability and utility in clinical settings.

### Loss function

Our research has contributed to refining the use of loss functions in medical image synthesis, focusing on the role of perceptual loss for enhancing the generation of fine structures in MR to CT synthesis within 2D and 2.5D setups. We explored various perceptual loss approaches, particularly examining the impact of different architectures and training datasets. This investigation led us to observe that perceptual loss models based on a broader range of natural images might offer more robust feature extraction capabilities than those trained exclusively on medical images [4]. Ultimately, our work offers insights into optimizing the use of perceptual loss for improving the quality of synthesized fine structures, providing a basis for further research in this area. For 3D setups, however, there is no perceptual loss available on natural images and only limited availability for perceptual loss trained on medical data.



**FIGURE 22.2** – Comparison of MR to CT synthesis using a traditional L1 loss and L1 loss combined with our structure-prioritized loss on a pre-trained airways segmentation model embeddings. Sagittal slices are presented, including a zoomed-in view of the bronchi reconstruction.

Expanding upon these findings, our most recent study introduced a structure-prioritized loss function, desi-

gned to enhance the detail and accuracy of specific anatomical features in synthesized images [5]. This approach, using feature maps from a pre-trained model on a specific segmentation task, demonstrated superior performance in reconstructing detailed structures like airways, beyond what traditional global reconstruction methods achieve. Our findings emphasize the potential of combining these innovative loss function to refine MR to CT image synthesis, aiming for both global image quality and precise structural detail. Future directions will explore the extension of these methods to other anatomical parts and further refine the synthesis process, ensuring the broader applicability and utility in clinical settings.

| Model                      | Dice          | MAE          |
|----------------------------|---------------|--------------|
| L1 loss                    | 0.5521        | 35.40        |
| L1 + percep. loss          | 0.5507        | 35.24        |
| <b>L1 + SP loss (ours)</b> | <b>0.5751</b> | <b>33.07</b> |

**TABLE 22.1** – Comparison of 3D networks performances with different loss functions, using MAE and airways Dice score between real CT and synthesized CT [5].

### Bibliographie

- [1] A. Longuefosse, B. Denis de Senneville, G. Dournes et al. MR to CT synthesis using GANs : a practical guide applied to thoracic imaging. International Conference on Information Visualization Theory and Applications, IVAPP (2024).
- [2] A. Longuefosse, J. Raoult, I. Benlala et al. Generating high-resolution synthetic CT from lung MRI with ultrashort echo times : initial evaluation in cystic fibrosis. Radiology, 308 (2023).
- [3] A. Longuefosse, B. Denis de Senneville, G. Dournes et al. Incorporating 3D information in 2.5D networks : strategies for MR to CT synthesis. IEEE 20th International Symposium on Biomedical Imaging, ISBI (2024).
- [4] A. Longuefosse, B. Denis de Senneville, G. Dournes et al. On the use of perceptual loss for fine structure generation : Illustration on lung MR to CT synthesis. IEEE 20th International Symposium on Biomedical Imaging, ISBI (2024).
- [5] A. Longuefosse, B. Denis de Senneville, G. Dournes et al. Fine Structure-Prioritized Loss for Enhanced MR to CT Translation. Submitted to Medical Imaging with Deep Learning, MIDL, (2024).



# MODELING APPEARANCE OF POROUS MATERIALS

Simon LUCAS

Direction : Pascal BARLA

Co-Direction : Romain PACANOWSKI

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

**P**orosity is present in a large number of materials. Natural materials such as soils and rocks have porosities of various kinds, ranging from aggregates formed by the cohesion of particles, to tubular pores created by the passage of water or by living beings, or cracks created by the action of various forces. Other non-natural materials also have pores, such as the cavities in the surface of ceramics. As porosity has a significant impact on the appearance of materials, it is vital to characterise it correctly in the context of computer graphics.

## Porous material analysis

Porous structures can be found at different scales. The smallest porosities are in the nanometre range. They are too small to be studied in a geometrical optics context, but can be taken into account in aggregate (micro) materials. Our work focuses on scales between the micron and the millimetre, on the BRDF (Bidirectional Reflectance Distribution Function). Above the millimetre, pores are large enough to be visible to the naked eye, and can therefore be modelled directly by geometric details.

Our main objective is to characterise the effects present in porous materials made up of opaque aggregates, both surface and volume effects. To do this, we use simulations to trace light paths in structures composed of clusters of opaque spheres (mono-disperse distribution). The clusters of spheres are parameterised by their porosity, or the proportion of void in the material, and by their anfractuosity, which determines whether they are compacted at the surface (cf. Figure 23.1).

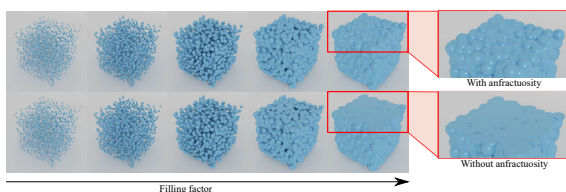


FIGURE 23.1 – Various porous structures used to compute light transport simulations.

Our simulations enable us to assess the relevance of existing porous material models. We show that these models can only approach simulations for restricted structures : the appearance space of porous materials remains largely uncovered. Our work highlights the effects that a general BRDF model should reproduce, which will have applications not only in computer graphics, but also in remote sensing for astronomy or ecology.

The appearance of objects is affected by their microscopic structure. Here we are interested in the appearance of porous materials. The presence of pores has been

modelled in two ways in existing work : by modifying the surface state of a material, or by modifying its volume properties. However, there is no model of porous materials that is generic enough to take into account both surface and volume effects. We propose a study of these effects using light transport simulations. This approach reveals the limitations of existing models called Bidirectional Reflectance Distribution Function (BRDF), as well as promising avenues for the future development of a material model that is sufficiently general to take into account variations in porosity and surface anfractuosity.

## Micrograins

We introduce a new BSDF model for the rendering of porous layers, as found on surfaces covered by dust, rust, dirt, or sprayed paint. Our approach is based on a distribution of elliptical opaque micrograins, extending the Trowbridge-Reitz (GGX) distribution [7, 4] to handle pores (i.e., spaces between micrograins). We use distance field statistics to derive the corresponding Normal Distribution Function (NDF) and Geometric Attenuation Factor (GAF), as well as a view- and light-dependent filling factor to blend between the porous and base layers. All the derived terms show excellent agreement when compared against numerical simulations.

Our approach has several advantages compared to previous work. First, it decouples structural and reflectance parameters, leading to an analytical single-scattering formula regardless of the choice of micrograin reflectance. Second, we show that the classical texture maps (albedo, roughness, etc) used for spatially-varying material parameters are easily retargeted to work with our model. Finally, the BRDF parameters of our model behave linearly, granting direct multi-scale rendering using classical mip mapping.

Porosity is common in natural and man-made materials. We distinguish two types of structures : granular materials such as sand or earth where many tiny grains are distributed in a volumetric fashion ; and porous layers such as dust or dirt where grains are distributed onto



FIGURE 23.2 – Micrograins on various surfaces.

another surface. The former type requires complex volumetric rendering techniques (e.g. [6]), or must be restricted to simpler configurations to yield suitable BRDF approximations (e.g., [9, 8]). In this paper, we focus instead on the latter type, which permits the blending of two materials (the porous layer on the base surface) with varying grazing-angle effects. Our objective is to derive a physically-based BSDF model for such material configurations.

The key idea of our approach is to consider a distribution of micrograins on a surface, from which we derive a BSDF model according to microfacet theory [5]. The micrograins are thus considered too small to be visible to the naked eye. Structures larger than a micrograin (e.g., meso-scale aggregates) should be modelled explicitly (e.g., through displacement maps).

Our contribution is twofold. First, we define a microfacet Normal Distribution Function (NDF) from the distribution of micrograins. This new NDF extends the Trowbridge-Reitz (GGX) distribution to handle porosity (i.e., the spaces between micrograins). Second, we define visibility terms, the visible filling factor of a porous layer on a base surface, as well as the Geometric Attenuation Factor (GAF) required by microfacet theory. The main advantage of relying on a micrograin distribution is that it decouples structural from reflectance parameters. As a result, our model is analytic for a range of micrograin reflectances (diffuse, diffuse + specular and conductor). Moreover, all material parameters are easily assigned to material texture maps, which eases the adoption of our model in existing asset creation workflows (see Figure 23.1).

## Bibliographie

- [1] Order statistics, David, Herbert A and Nagaraja, Haikady N, 2004, John Wiley & Sons
- [2] The Malia Rendering Framework, Arthur Dufay and David Murray, and Romain Pacanowski, and others, <https://gitlab.com/mrf-devteam/mrf/main.md.html>, 2021
- [3] Microfacet theory for non-uniform heightfields, d'Eon, Eugene and Bitterli, Benedikt and Weidlich, Andrea and Zeltner, Tizian, 2023, Association for Computing Machinery
- [4] B. Walter and S. R. Marschner and H. Li and K. E. Torrance, Microfacet Models for Refraction Through Rough Surfaces, Computer Graphics Forum, EGSR proceedings, 2007
- [5] K. E. Torrance and E. M. Sparrow, Theory for Off-Specular Reflection From Roughened Surfaces\*, J. Opt. Soc. Am., 1967
- [6] Müller, Thomas and Papas, Marios and Gross, Markus and Jarosz, Wojciech and Novák, Jan, Efficient Rendering of Heterogeneous Polydisperse Granular Media, "ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)", 2016
- [7] T. S. Trowbridge and K. P. Reitz, Average irregularity representation of a rough surface for ray reflection, J. Opt. Soc. Am., 1975,
- [8] , d'Eon, Eugene, An analytic BRDF for materials with spherical Lambertian scatterers, Computer Graphics Forum
- [9] Hapke, Bruce, Theory of Reflectance and Emission Spectroscopy 2012

# INTERACTIONS TANGIBLES POUR L'EXPLOITATION FERROVAIRE DANS LES CENTRES OPÉRATIONNELS

**Maudeline MARLIER**

**Direction : Arnaud PROUZEAU**

**Co-Direction : Martin HACHET**

**Encadrement : Nicolas RENOIR (SNCF)**

*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

---

**W**ith the growth of data and digital tools, the development of artificial intelligence and the goal of increasing train traffic, the role of humans will evolve. Today, operators are responsible for monitoring and managing railway traffic. The goal of this thesis is to test if tangible interactions can help operators in decision-making during incidents. For that, we conducted an in-the-field study to identify the limits of the current approaches in control centres. To test our hypothesis, we did a formative study by developing a prototype to investigate the use of actuated tangible tokens. Our results suggest that actuated tangible tokens are promising for collaboration, allowing the addition of information with a simple gesture and understanding easily through movements on the table. Future work will measure the impact of actuated tangible interactions on a focused attention task. An initial study will show whether actuated tangible tokens reduce inattention blindness compared with a simple projection on the table. Then, we will go further and compare actuated tangible tokens and projection during a remote collaboration task.

---

## Introduction

In France, railway traffic is monitored and managed by operators in control centres. During disruptions, operators must minimise the repercussions on the entire traffic flow and ensure quicker response times to inform passengers. However, incident management currently requires operators to take time to understand the situation, to have an overall view of the traffic and its particularities at a given moment in order to be able to make a decision. Our research hypothesis in this thesis is that integrating a table in the centre of the room with tangible tokens could help operators in decision-making by improving collaboration, communication and resolution solutions visualisation.

Ishii and Ulmer [1] define the Tangible User Interface (TUI) as an interface that "gives physical form to digital information, employing physical artefacts both as representations and controls for computational media". Tangible interactions assist in structuring spatial information and help when communication is difficult with words. Tangible interactions have been also studied specifically with multi-user collaboration [2]. TUI seems to have a remarkable effect on engagement during collaborative decision-making task. Specifically, they can facilitate an understanding of complex systems while supporting collaboration.

### **Control centres : field study**

We visited 7 different control centres in Paris and Bordeaux. These visits allowed us to understand better how operators are organised and how they manage both routine operations and critical events. We observed that during an incident, operators are tasked with adjusting

services, devising alternative strategies, coordinating resource availability and informing passengers. All of these while still monitoring the traffic on the rest of the network. For that, operators must possess a comprehensive grasp of the network's conditions at all times, along with a high level of situational awareness. With the current tools, information is scattered across all tools or in operator's head if they received a call for example. Information is not shared with the entire room and it can be difficult to get an overall picture. Operators are forced to store information in their working memory and to mobilise cognitive resources for mental representations. This leads to a slowdown in decision-making and limits collaboration.

With all these observations, we concluded that operators have some difficulties sharing information or maintaining an appropriate level of awareness of the situation when disruption happens on the network.

### **Prototyping tangible interactions for decision-making in control centres**

We explored a solution that takes benefits from both a shared display and physical objects that can be manipulated. The goal is to leverage collaboration and situation awareness. For that, we propose to explore the use of a digital tabletop with actuated tangible tokens (i.e. tokens that can move by themselves on the table and be moved by operators) as a tool for railway control, illustrated in Figure 24.1.

To demonstrate a concrete use of our prototype, we draw up a scenario inspired by a real event with an incident-management expert. In this scenario, operators can visualise, create, compare or delete incident

resolution solutions. For example, a train is stopped in the middle of the tracks between two stations. The operator can double-tap on the token that will then demonstrate on the network the visualisation of the solution (the next train will push the stopped train) illustrated in Figure 24.2.

As such a system would represent an important change in operators' work environment and work method. We wanted to first assess their acceptance of the use of actuated tangible interactions and explore with them how it could be integrated with their current practices and setup. Therefore, we developed a prototype and performed expert feedback sessions with 37 operators including 34 in a control centre in charge of the high-speed trains in the west of France and three crisis management experts. Our results suggest that such a technology could help operators during crises with information sharing, context understanding and time-saving.

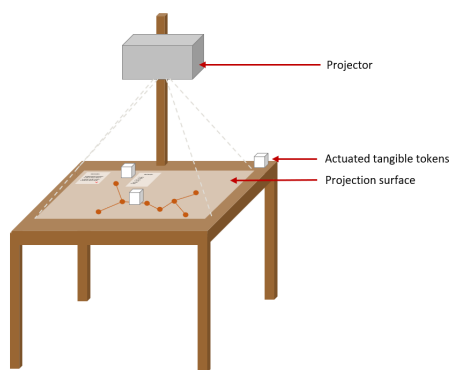


FIGURE 24.1 – Schema of the setup of the prototype that we implemented composed of a projector (for top-down projection) and actuated tangible tokens able to roll on a table.

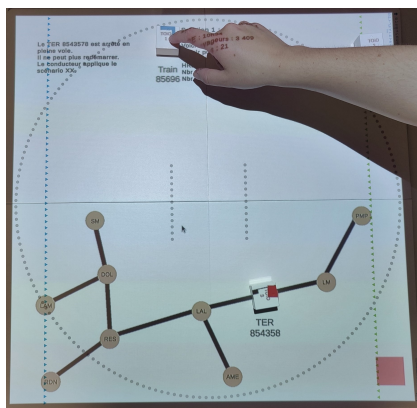


FIGURE 24.2 – Photographs of an interaction on the prototype : operators double taps on the token to play the solution

### Actuated tangible tokens as output for collaboration

As explained before, operators need to maintain an appropriate level of awareness of the situation. Several times, operators missed information because they did not see the app notification or did not hear oral information. This phenomenon in control centres is called

"inattentional blindness". It occurs when an individual is blind to the presence of an entire object while performing a distracting task [3].

This second project aims to study remote collaboration between two centres. We formulate the hypothesis that actuated tangible tokens for feedback during incident management could be better than current visual information on screen in the context of remote collaboration. In order to test our hypothesis, we will conduct two experiments : (i) With an abstract task, we will compare inattentional blindness with actuated tangible tokens and with visual-only feedback both on a table with a single user. (ii) We will create a more specific task for remote collaboration. Two participants will use their own tangible device. To compare actuated tangible tokens and visual feedback we will create two different ways to communicate apart from verbal communication. The tasks and measures for both experiments have yet to be defined.

## Conclusion

Tangible interactions have been used in different domains, but have not been studied in incident management situations for railway traffic. This thesis aims to study these human-machine interactions, taking a global approach by creating a prototype and then studying more specifically actuated tangible tokens for remote collaboration.

## Bibliographie

- [1] Brygg Ullmer and Hiroshi Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM systems journal* 39, 3.4, 915-931.
- [2] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable : exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. 139-146.
- [3] Jonathan E Attwood, Christopher Kennard, Jim Harris, Glyn Humphreys, and Chrystalina A Antoniadis. 2018. A comparison of change blindness in real-world and on-screen viewing of museum artefacts. *Frontiers in Psychology* 9, 151.



# VÉRIFICATION ET SYNTHÈSE DE SYSTÈMES DISTRIBUÉS À SYNCHRONISATION FAIBLE OU COMMENT APPLIQUER LA THÉORIE DES AUTOMATES À LA VÉRIFICATION DE SYSTÈMES DISTRIBUÉS

**Corto MASCLE**

**Direction : Anca MUSCHOLL**

**Co-Direction : Igor WALUKIEWICZ.**

*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

Les programmes distribués sont une source de difficulté pour la vérification. La composition de systèmes parallèles qui peuvent se synchroniser a souvent la puissance du produit de ces systèmes, rendant la vérification bien plus complexe. Dans le cadre de la synthèse, à ces problèmes s'ajoute celui de la propagation de l'information, car les actions d'une composante ne peuvent tenir compte que de la configuration globale. Dans cette thèse, nous étudions des systèmes distribués avec une communication très restreinte entre les composantes : d'une part les systèmes où la communication se fait via la prise et le relâchement de locks, et d'autre part ceux où les programmes communiquent par des variables partagées. Nous établissons la décidabilité de la vérification et de la synthèse dans plusieurs cas, et contrastons ces résultats avec des résultats d'indécidabilité. Le résumé qui suit explore l'application de divers types d'automates à l'analyse formelle de systèmes distribués.

## Deadlocks et $\omega$ -automates à piles

**Deadlocks.** Un *deadlock* est une situation où plusieurs agents se bloquent parce que chacun attend une action des autres pour agir. Un exemple de deadlock est si votre fournisseur d'internet vous envoie un mail vous demandant de confirmer votre abonnement : Vous attendez d'avoir accès à vos mail pour confirmer, tandis qu'il attend votre confirmation pour vous donner l'accès internet. En informatique, ce genre de problème peut notamment arriver dans les systèmes distribués lorsqu'ils partagent des *locks*. Un *lock* est un marqueur utilisé pour contrôler l'accès à une ressource ou une variable partagée. Lorsqu'un agent veut utiliser cette ressource, il *prend le lock* associé, ce qui empêche les autres agents d'y accéder tant qu'il ne l'a pas *rendu*.

**Modélisation.** On utilise la convention, commune en vérification formelle, que les agents peuvent se modéliser par des systèmes à états finis avec une pile. La pile sert à modéliser des appels récursifs du programme, et les états servent à suivre l'endroit du programme qui est exécuté, ainsi qu'une quantité bornée de mémoire (par exemple la valeur de variables booléennes). On surapproxime le comportement de l'agent par un tel système : si aucune exécution du vrai système ne rencontre d'erreur, alors ce sera aussi le cas pour le système abstrait. Cependant, il se peut que le second ait des erreurs que le premier n'a pas : c'est le prix à payer pour des problèmes de vérification décidables.

**Vérification et synthèse** Les problèmes considérés dans ma thèse se divisent en deux catégories : les pro-

blèmes de vérifications et ceux qui creusent et les problèmes de synthèse. Dans les premiers, on demande si un système donné peut avoir un certain comportement indésirable, tandis que dans les deuxièmes, le système a la possibilité de choisir certaines actions, et on demande s'il existe une stratégie pour éviter ces comportements.

**Mais attention**, comme on s'intéresse à des systèmes distribués, ces stratégies doivent pouvoir être implémentées de manière locale : Pour prendre ses décisions, un processus ne peut pas tenir compte de ce que les autres processus ont fait.

## Systèmes à locks

Un premier modèle consiste à considérer un nombre fixé de processus prenant et relâchant des locks, chacun modélisé par un automate à pile [4]. Ce modèle donne des problèmes indécidables en général, mais on peut contraindre les échanges de locks pour obtenir des algorithmes.

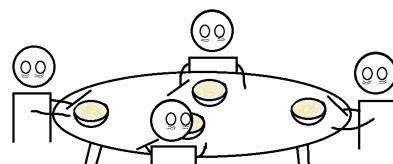


FIGURE 25.1 – Le dîner des philosophes au moment du drame.

Notamment, on peut restreindre chaque processus à utiliser au plus deux locks (comme dans le célèbre dîner des philosophes), ou encore demander que les locks



soient gérés comme dans une pile (un processus ne peut relâcher que le lock pris le plus récemment). Ces systèmes sont appelés LSS (Lock-Sharing Systems).

Pour analyser ces systèmes, on montre que les projections d'une exécution sur chaque processus peuvent être résumées en de petites descriptions contenant toute l'information nécessaire à la synchronisation.

Nous pouvons ainsi encoder les contraintes de locks comme des langages d'automates. Les propriétés à vérifier sont, elles, écrites comme des automates sur des runs infinis, ou  $\omega$ -automates. En effet, on veut pouvoir exprimer des choses comme "le système peut s'exécuter indéfiniment sans blocage". Comme les systèmes sont modélisés par des automates à piles, on obtient des  **$\omega$ -automates à piles**.

**Théorème 1** ([2, 6]). *La vérification de propriétés régulières locales sur les LSS est NP-complète.*

Les mêmes techniques s'appliquent dans le cas de la synthèse :

**Théorème 2** ([2, 6]). *L'existence d'une stratégie garantissant une propriété régulière locale est  $\Sigma_2^P$ -complète.*

## Vers les arbres infinis

Forts des techniques utilisés dans la partie précédente, nous nous sommes demandé si ces résultats pouvaient être étendus au cas où les processus peuvent créer de nouveaux locks et d'autres processus pendant l'exécution. La réponse est OUI, EXCELLENTE IDÉE!!! Ces systèmes sont alors appelés Dynamic LSS, ou DLSS. On fait encore une fois l'hypothèse que les locks sont gérés comme dans une pile.

On remplace les exécutions linéaires d'avant par des arbres : Un processus qui s'exécute correspond à une branche. Chaque fois qu'un processus en crée un autre, cela génère un noeud à droite, qui correspond au début de l'exécution du processus créé [5].

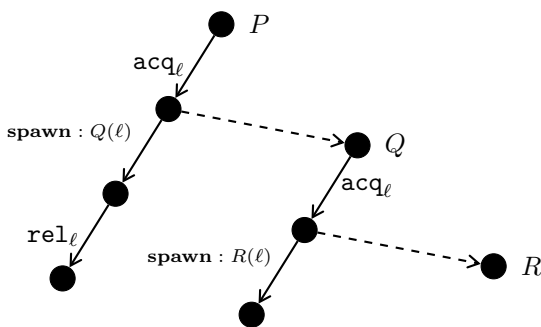


FIGURE 25.2 – Représentation (simplifiée) d'exécution d'un DLSS : le lock  $\ell$  est partagé, le processus  $P$  crée  $Q$  qui crée  $R$ .

Nous avons créé un algorithme qui, à partir d'un DLSS, construit un **automate d'arbres** qui reconnaît les arbres correspondant à ses exécutions. La principale difficulté est de s'assurer qu'il n'y a pas de contradiction entre les différentes branches.

**Théorème 3** ([7]). *La vérification régulière de DLSS est EXPTIME-complète, et polynomiale si le nombre de locks par processus est fixé.*

## Des variables partagées aux automates à registre

Passons à une mode de communication un peu différent. Imaginons un grand nombre de processus identiques communiquant via une variable partagée  $x$ . Les processus peuvent lire la valeur de la variable, ou bien y écrire une nouvelle valeur. De plus, les messages sont signés, c'est-à-dire qu'avec la valeur de la variable est écrit l'identifiant du processus qui l'a écrite.

On considère le nombre de processus comme un paramètre qui peut être arbitrairement grand. Un problème classique sur ce genre de modèles est la *couverture* : "Est-ce l'état  $q_{\text{erreur}}$  peut être atteint à partir d'un certain nombre de processus ?".

Une difficulté est le fait que les identifiants inscrits dans la variables ne sont pas bornés : les automates habituels, à alphabet fini, ne sont pas assez expressifs ! Pour cette raison, on se tourne vers des **automates à registres**. Ceux-ci permettent de lire des mots sur des alphabets infinis, en utilisant des registres pour stocker les valeurs reçues. Leurs transitions contiennent des tests du type "la valeur reçue est égale à celle du 3e registre".

Le modèle ci-dessus s'appelle les BNRA (pour Broadcast Networks of Register Automata) [1]. Une des contributions de ma thèse est d'avoir montré que la couverture est décidable sur les BNRA.

**Théorème 4** ([3]). *Le problème de couverture est décidable sur les BNRA.*

Ce théorème vient avec une caractérisation de la complexité du problème en fonction du nombre de registres par processus, l'adaptations à plusieurs extensions du modèle, et bien d'autres choses encore !

## Bibliographie

- [1] Delzanno, G. and Sangnier, A., and Traverso, R. Parameterized verification of broadcast networks of register automata : RP 2013.
- [2] Gimbert, H. and Mascle, C. and Muscholl, A. and Walukiewicz, I. Distributed controller synthesis for deadlock avoidance : ICALP 2022.
- [3] Guillo, L. and Mascle, C. and Waldburger, N. Parameterized Broadcast Networks with Registers : from NP to the Frontiers of Decidability : FoSSaCS 2024.
- [4] Kahlon, V. and Ivančić, F. and Gupta, A. Reasoning about threads communicating via locks. CAV 2005.
- [5] Lammich, P. and Müller-Olm, M. and Seidl, H. and Wenner, A. Contextual locking for dynamic pushdown networks : SAS 2013.
- [6] Mascle, C. Model-checking lock-sharing systems against regular constraints : Arxiv preprint, 2022.
- [7] Mascle, C. and Muscholl, A. and Walukiewicz, I. Model-checking parametric lock-sharing systems against regular constraints : CONCUR 2023.

# PROGRAM SYNTHESIS : SCALING ENUMERATIVE APPROACHES

Théo MATRICON

Direction : Nathanael Fijalkow

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Program synthesis is a task first imagined by Church that is still hailed as a holy grail. The goal of generating a correct program matching a specified pairs of inputs, outputs examples has been tackled in many different ways. We will first look at the enumerative approaches that can solve such tasks in a domain agnostic manner. Then we will show how we contributed to improving such algorithms and how far the state of such algorithms have changed. In order to scale, we will also show how to leverage semantic properties automatically to improve scaling of our approaches. We will also look at including knowledge of the outside world in programs.

## Program Synthesis From Examples

Program synthesis from examples is the following task : given a grammar  $G$  that describes the space of possible programs, and pairs of inputs, outputs examples find a program that matches the given examples. Let us show how this could work with an example :

```
f([7, 8, 9]) = [8, 8, 10]
f([15, 4, 8, 9]) = [4, 8, 10, 16]
f([]) = []
```

and here is a possible solution :

```
def f(x):
    g = lambda y: y + 1 if isOdd(y) else y
    x = map(g, x)
    x = sorted(x)
    return x
```

There are different approaches to tackle this synthesis task. The first one being the enumerative approach [2], the idea is that we are given some cost function over programs and we will enumerate each program by order of increasing cost. Then each program is tested against the input output examples. The advantage of this approach is that it requires minimal knowledge and it is domain agnostic. A second approach is counter example guided synthesis (CEGIS) [1], this assumes that each concept of the grammar is translated into typed first order logic, this becomes a theory. As before, programs are enumerated, an enumerated program is rejected if it does not match the logical specification, if it matches then it is sent to an SMT solver which guarantees that the program matches the examples or provides a counter example in the form of first order logic that is added to the specification. The advantage of such a method is that it enables to prune large part of the search space however it requires the specification of the concepts and that such concepts can be embedded into first order logic. A third approach which has boomed recently is using neural networks to produce programs. Here code

is no different from text, thus it has skyrocketed with the rise of large language models, however radically different from the previous approaches is the fact that the output program has no guarantee of being correct on the given examples, it even does not guarantee that the program is well typed nor syntactically correct.

## Enumerative Approach

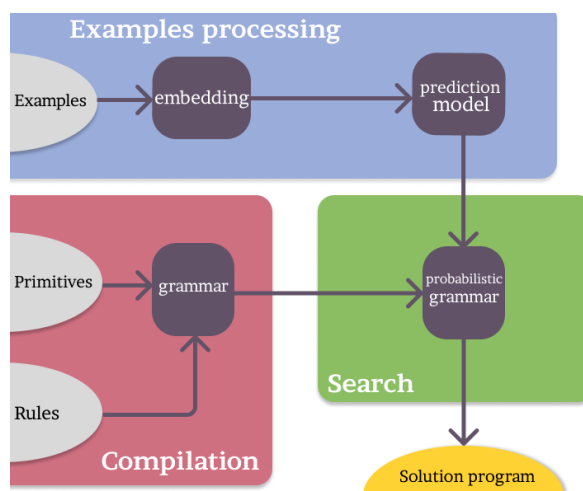


FIGURE 26.1 – Pipeline of a classic enumerative program synthesis solver

First, let us describe the classic enumerative program synthesis pipeline as shown on Figure 26.1. Examples are fed into a prediction model, usually a neural network, which produces probability distributions at the non terminal level of the grammar, this will be used as a cost function. Users only have to specify the syntax and the semantic of the primitives such as texttsorted for example. This approach relies on two key components : the quality of the cost function with respect to the task and the speed of the enumerative algorithm. The faster the enumerative algorithm the less impactful the cost function is, therefore to support a wide variety

of tasks it is necessary to have efficient enumerative algorithms. Our first contribution was the HeapSearch algorithm [4], an enumerative algorithm that satisfies the property of enumerating programs deterministically in the correct order while being twice as fast as previous approaches. Furthermore, since the future is a parallel one, we also contributed with an algorithm that enables a linear scaling with the number of processors by splitting the probabilistic grammar. A forthcoming contribution is a new algorithm that is an order of magnitude faster than HeapSearch and other algorithms and scales dramatically better than them with the increase of the grammar complexity.

Another way to reduce the space of programs searched is to look into semantic equations for example  $\text{sorted}(x) == \text{sorted}(\text{sorted}(x))$  or  $a + b == b + a$ . Such equations can be used to reject programs that are semantically equivalent to another during the enumeration since enumerating is rather cheap compared to testing a program on the examples. In order to encode such constraints we can use deterministic bottom up tree automata since we can leverage the fact that our enumerative approaches are bottom up this naturally flows with the automaton. In order not to have the user specify a possibly humongous list of equations, we can start at no equation, then enumerate small programs and evaluate them on a wide range of examples and build set of equivalent programs. This produces a set of equivalent programs used as equations. A nice property is that this can be done once and for all at the language level, only the automaton needs to be saved.

### Knowledge powered programs

Classic program synthesis is limited in the sense that programs cannot talk about the real world. Our contribution [5] is a technique that enables solving such tasks, without relying on machine learning that is the knowledge manipulated is certain there is no possible hallucination. An example of a knowledge powered task is on Figure 26.2. The task is to go from a city to its country and its phone code. In order to tackle such a task that is actually quite common in the real world, we will make use of a knowledge graph. The idea is to split the task in subtasks, on the example we split around "code:" since this part is constant and then we try to solve each subtask. To solve a subtask we try in parallel to solve it using the classical program synthesis pipeline that does not talk about knowledge and at the same time we try to find a common path amongst the examples on the knowledge graph that can solve the subtask, this translates into a SPARQL query. On the example to get from texttt"France" to texttt"Paris" you have to follow the CapitalOf edge in the graph.

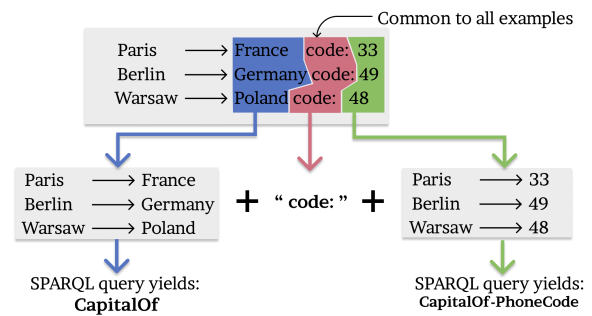


FIGURE 26.2 – Overview of the techniques used to deal and solve a knowledge powered program synthesis task

## Conclusion

Our contributions have paved the way for faster and more complex program synthesis approaches. We have contributed with faster and better enumerative algorithms for enumerative approaches but they can also be used by CEGIS approaches since they also need enumerative approaches. We also looked into including real knowledge in programs and contributed with an algorithm so solve them efficiently. However, CEGIS approaches and LLM have advantages that cannot be underestimated and an avenue of future work is combining these approaches in order to leverage the advantages of the different approaches.

## Bibliographie

- [1] Alur, Rajeev, et al. Syntax-guided synthesis. IEEE, 2013.
- [2] Balog, Matej, et al. "Deepcoder : Learning to write programs." arXiv preprint arXiv :1611.01989 (2016).
- [3] Li, Yujia, et al. "Competition-level code generation with alphacode." Science 378.6624 (2022) : 1092-1097.
- [4] Fijalkow, Nathanaël, et al. "Scaling neural program synthesis with distribution-based search." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 6. 2022.
- [5] Matricon, Théo, Nathanaël Fijalkow, and Gaëtan Margueritte. "WikiCoder : Learning to Write Knowledge-Powered Code." International Symposium on Model Checking Software. Cham : Springer Nature Switzerland, 2023.

# USING SEQUENCE ANALYSIS TO CHARACTERIZE BEHAVIORS DURING A CREATIVE PROBLEM SOLVING TASK

**Axel PALAUDE**

**Direction : Thierry VIEVILLE**

**Co-Direction : Frédéric ALEXANDRE**

*Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

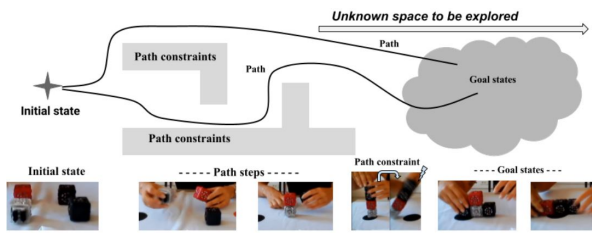
---

**C**reative problem solving (CPS) activities observation can enrich our understanding of human learning by facilitating the access to adaptive behaviors related to learning mechanisms. Actions taken by an individual during a CPS task are situated in time, thus allowing to create time sequences of observations. Sequence analysis should allow us to overcome the limits of current observations of such activities. We aim to develop a novel analysis approach on a set of time sequences based on the configurations of a game-based robotic activity consisting of assembling modular robotic cubes. This allows the creation of clusters of sequences or subsequences of actions, in order to help the characterization of different types of behaviors. We base our work on CPS configuration identification analyzed in Romero et al. (2018) [1] on the CPS activity named CreaCube in order to characterize a subject's individual task. Our clustering method requires to measure the similarity between sequences, thus calling for a definition of a distance between observables. We propose to consider a distance using Dynamic Time Warping as a measure of similarity. By having access of a similarity, we compare the results of multiple executions of clustering methods to existing models of behaviors from the literature.

---

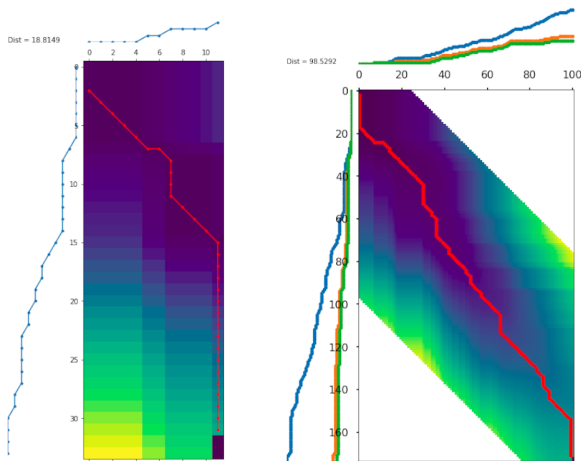
While research in well-defined problem solving has permitted to characterize certain tasks such as the Tower of Hanoi and to formalize general procedures for problem solving using computational representations (Newell & Simon, 1976) [2], the study of problem solving in learning sciences has broaden to include complex and domain-specific problem-solving tasks. In CPS tasks, a learner has to build representations of those problems, and use them to decide on a behavior to adopt, including goal choice. Actions undertaken by the learner may result in a change in the CPS situation that can be used as a feedback to evaluate the advancement of an intermediate situation and decide to adapt the behavior, according to their updated representation of the problem. We focus on CPS tasks in which the subject needs to search for a novel solution for ill-defined problems. The research objective is to characterize individuals' behaviors by identifying recurring behavioral patterns across experiments using clustering methods, and analyzing the resulting clusters with regard to categorizations of behaviors proposed in CPS research. We thus focus on a task named CreaCube, an ill-defined robotic game-based activity studied by Romero et al. (2018) [1]. In this CPS task, the participant is invited to manipulate four cubes of different colors. Each cube has different properties such as wheels or sensors, but these technological affordances are unknown to the subject when starting the activity. The goal of this CPS task is to build a vehicle composed of four pieces that can move

autonomously from one point to another, without any further initial information about the problem. Based on a subset of the CreaCube corpus, selecting only children (7-12 y.o) who performed the task individually (93 subjects), we have for each experiment a data set including a time sequence starting when the subject hears the instructions for the first time and stopping when the subject succeeds or gives up. In each sequence, some key events that can be observed have been annotated, resulting in 83 different types of observables annotated across all experiments. These observables, detailed in Mercier et al. (2022) [3], mainly correspond to the discovery of affordances (the discovery of properties of some cubes), the realization of the most common four-cubes structures, the criteria those structures fulfill (or not) with regard to the guideline, and the final outcome of the activity. Those time sequences are thus symbolic trajectories (see Figure 27.1), and we want to cluster those trajectories.



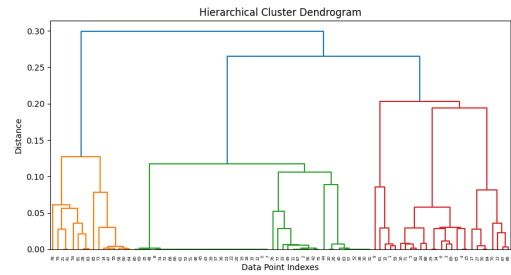
**FIGURE 27.1** – An abstract symbolic representation of a problem space and two possible trajectories that can solve it with Creacube states.

We propose to apply a K-nearest-neighbors (KNN) algorithm to this corpus of time sequences. This clustering algorithm requires the definition of a distance between trajectories. For this purpose, we use the Dynamic Time Warping (DTW) algorithm, since it allows to measure the similarity between two time sequences while slightly realigning them to mitigate small time variations among trajectories. In order to compute an average sequence from a set of trajectories (required for applying KNN), we transform symbolic trajectories into numerical sequences based on the increasing number overtime of Creacube 4-cubes structures based on the DT components of Guilford : fluidity, flexibility and innovation [4] (see Figure 27.2).



**FIGURE 27.2** – DTW distance on two sequences (left) and on two Creacube trajectories (right). The Creacube trajectories are composed of fluidity (number of structures done, blue), flexibility (number of different structures done, orange) and innovation (number of structures done by less than 5 % of subjects, green)

Taking advantage of our small dataset, we are introducing a method called *metaclustering* [5], consisting on creating a distance matrix storing the number of times two different sequences were clustered together with every possible clustering outcome. Divided by the total number of clustering executions, we can the apply hierarchical clustering in order to create a dendrogram allowing to see the proximity between each sequence (see Figure 27.3).



**FIGURE 27.3** – Dendrogram of the hierarchical clustering based on the metaclustering matrix.

We are comparing such grouping to existing behavioral models in psychology or educational science and to generalize this method to other problem-solving activities.

## Bibliographie

- [1] Romero, M., David, D., & Lille, B. (2018, December). CreaCube, a playful activity with modular robotics. In International Conference on Games and Learning Alliance (pp. 397-405). Springer, Cham.
- [2] Newel, A., & Simon, H. A. (1976). Computer science as empirical inquiry : Symbols and search. Communications of the ACM, 19(3), 113-126.
- [3] Mercier, C. (2022). An Ontology to Formalize a Creative Problem Solving Activity. IEEE Transactions on Cognitive and Developmental Systems.
- [4] Leroy, A., Romero, M., & Cassone, L. (2021). Interactivity and materiality matter in creativity : educational robotics for the assessment of divergent thinking. Interactive Learning Environments, 1-12.
- [5] Monti, Stefano et al. Consensus Clustering : A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning 52 (2003) : 91-118.



# OPTIMIZATION OF TASK-BASED APPLICATIONS ON HETEROGENEOUS COMPUTING NODES

Hayfa TAYEB

Direction : Abdou GUERMOUCHE

Co-Direction : Béranger BRAMAS

Encadrement : Mathieu FAVERGE

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

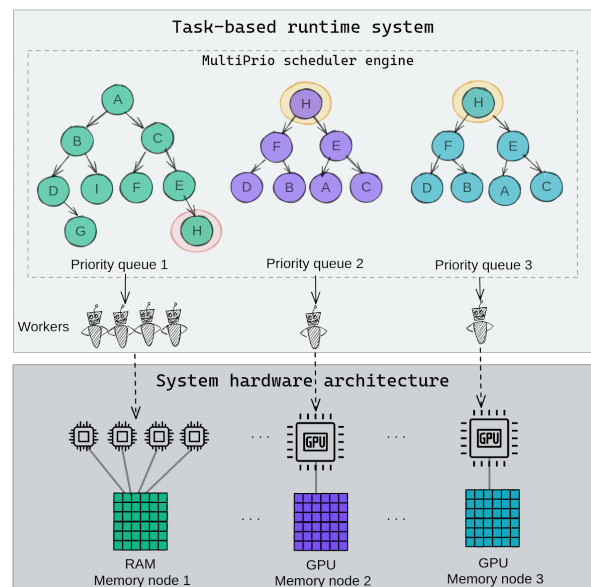
**E**fficient use of heterogeneous computing systems is essential for scientific and industrial applications. Task-based programming is effective, but scheduling is crucial for performance, especially on heterogeneous systems. We present MultiPrio, a dynamic task scheduler that aims to minimize completion time by considering resource affinity, task criticality, and data locality. Priority queues and heuristics are used to assign tasks to resources. Evaluation on Fast Multipole Method, dense and sparse linear algebra applications shows up to 31% performance improvement over state-of-the-art schedulers.

## Motivation

Modern high-performance computing systems are becoming increasingly heterogeneous, with a wide variety of parallelism, such as multiprocessors and accelerators, such as graphical processing units (GPUs). This complexity poses a challenge to application developers who want to take full advantage of these platforms. To simplify utilization of these architectures, task-based programming paradigms have emerged, representing applications as directed acyclic graphs (DAGs) of tasks with data dependencies. The execution of these tasks is managed by runtime systems, which serve as intermediaries between the user application and the underlying hardware infrastructure, orchestrating task execution and data movement across heterogeneous resources to optimize performance. Runtime systems like StarPU [1] employ dynamic scheduling strategies categorized into resource-centric (e.g., work stealing) and task-centric (e.g., Heterogeneous Earliest-Finish-Time) approaches. However, existing schedulers have limitations, particularly in scheduling irregular DAGs, and often consider only a limited set of criteria for decision-making. In response, we propose a multi-criteria scheduler that combines elements from both families, leveraging task-resource affinity and criticality scores for efficient workload distribution.

## Dynamic Scheduler Design

A heterogeneous computing system consists of different types of processing units, such as CPUs and GPUs. The system includes a set of memory nodes. In our study, we see the main RAM of a computing node as a single memory node despite the NUMA effects but otherwise the approach remains valid. A memory node can be either the main RAM, a GPU-embedded memory or disk memory.



**FIGURE 28.1** – MultiPrio scheduler example with 2 GPUs and 1 CPU. The tasks in the priority queues are ready for execution and are sorted according to their scores, which can be different depending on the type of processing unit. For example, the ready task *H* has the highest score on the GPU (and is stored at the top of the queue) and a lower score on the CPU.

Our proposed dynamic scheduler manages priority queues, each assigned to a memory node. Given the limited number of memory nodes in typical systems, the overhead remains reasonable. The scheduler inserts each ready task *t* in all respective priority queues such that *t* can be executed on the corresponding processing unit. Tasks are then duplicated in the data structures, since a task may have multiple implementations per resource. Each task is assigned a pair of scores, first an affinity score and then a criticality score. They are calculated using various heuristics. Heuristics use estimation of task execution time which is provided by a history-

based performance model from the runtime system. The scores are used to sort tasks without immediately deciding which worker is responsible for executing them. Idle workers select prioritized tasks with the most available data from the memory node's queue. This strategy allows slower workers to take over tasks when faster workers are too busy, under certain conditions. The aim is to balance the workload on the heterogeneous resources while prioritizing critical and local tasks for the fastest ones.

## Experimental Evaluation

We evaluate our proposed scheduling strategy on a heterogeneous computing node, studying the performance of regular and irregular task-based applications. This covers various real-world scenarios with different application characteristics. Regular applications have task graphs that remain the same regardless of the input data, like Cholesky factorization in dense linear algebra. User priorities for such applications are optimized offline by experts. It is also possible to define a fine-grained task submission strategy for the DAG. That's the case of several routines in the CHAMELEON library. On the other hand, irregular applications, have task graphs that change with the input data. These applications, like the Fast Multipole Method (FMM), which is used in various scientific simulations, and QR\_MUMPS, a sparse linear algebra library, pose more challenges for schedulers. As an illustration, we show the performance results of QR multi-frontal factorization from QR\_MUMPS on a set of sparse matrices detailed in the table 28.1. It introduces highly irregular workloads, including tasks of different granularities, while exhibiting variable memory consumption. In this evaluation, we compare the MultiPrio scheduler and its potential with the existing state-of-the-art schedulers (Dmdas [2] and HeteroPrio [3]) in StarPU on a heterogeneous node with 2 Nvidia V100 GPUs and 2 Intel Xeon Gold with 16 cores each. The figure 28.2 shows the performance ratio of each scheduler in comparison to the Dmdas scheduler, used as a reference. Higher ratios indicate better results, i.e. shorter execution times. MultiPrio optimizes the performance of most matrices, with an average gain of 31% over Dmdas. This is mainly attributed to the increased parallelism in the DAG.

| Matrix name  | Rows    | Cols   | op.count<br>(Gflop) |
|--------------|---------|--------|---------------------|
| cat_ears_4_4 | 19020   | 44448  | 236                 |
| flower_7_4   | 27693   | 67593  | 889                 |
| e18          | 24617   | 38602  | 1439                |
| flower_8_4   | 55081   | 125361 | 3072                |
| Rucci1       | 1977885 | 109900 | 5527                |
| TF17         | 38132   | 48630  | 15787               |
| neos2        | 132568  | 134128 | 31018               |
| GL7d24       | 21074   | 105054 | 26825               |
| TF18         | 95368   | 123867 | 229042              |
| mk13-b5      | 135135  | 270270 | 352413              |

TABLE 28.1 – Matrices used for QR\_MUMPS sorted by Gflops count.

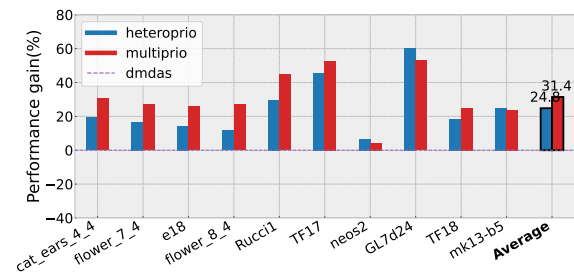


FIGURE 28.2 – Performance of QR factorization (ordering METIS) using 2 GPUs and 32 cores, relative to the Dmdas scheduler. Matrices sorted by Gflops count.

MultiPrio is efficient in optimizing irregular task-based applications and remains comparable for regular ones. In CHAMELEON, task criticality is the primary factor affecting performance, followed by resource management. Dmdas with user-defined priorities is able to further optimize performance by leveraging user expertise on regular DAGs. Additionally, optimizing data transfer times is critical for such intensive workloads.

## Conclusion

Effective use of heterogeneous computing systems for resource-intensive applications depends on efficient scheduling. We propose a novel multi-criteria dynamic scheduler that finds a balance between resource affinity, task criticality, and data locality. The novelty of our scheduler lies in its ability to automatically and efficiently schedule irregular applications, while being competitive with highly tuned schedulers on a more regular workload. Further refinement of the data locality strategy could be an area for improvement. Another interesting research direction, we aim to extend our research to another paradigm with recursive tasks. These tasks submit subgraphs at runtime. This strategy exposes different task sizes in the DAG, providing a sufficient amount of large-granularity tasks to efficiently utilize GPUs, along with fine-granularity tasks to take advantage of CPUs and thus unlock more parallelism.

## Bibliographie

- [1] Augonnet, C., Thibault, S., Namyst, R. & Wacrenier, P. StarPU : a unified platform for task scheduling on heterogeneous multicore architectures. *Concurrency And Computation : Practice And Experience*. **23**, 187-198 (2011)
- [2] Augonnet, C., Clet-Ortega, J., Thibault, S. & Namyst, R. Data-Aware Task Scheduling on Multi-accelerator Based Platforms. *2010 IEEE 16th International Conference On Parallel And Distributed Systems*. pp. 291-298 (2010)
- [3] Agullo, E., Bramas, B., Coulaud, O., Darve, E., Messner, M. & Takahashi, T. Task-based FMM for heterogeneous architectures. *Concurrency And Computation : Practice And Experience*. **28**, 2608-2629 (2016)

# IMPROVING MOTOR REHABILITATION AFTER A STROKE WITH ARTIFICIAL INTELLIGENCE AND NON-INVASIVE NEUROTECHNOLOGIES

David TROCELLIER

Direction : Fabien LOTTE

Co-Direction : Bernard N'KAOUA

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Using Brain computer interfaces (BCI) for stroke therapy has shown significant improvements in the movement abilities of stroke patients compared to traditional therapies. This effectiveness is attributed to the therapy's ability to establish a proprioceptive loop, where sensory feedback is directly linked to brain activity. This simultaneous activation helps to rebuild stronger connections in the brain based on neural plasticity, or the brain's ability to adapt and reorganize itself. During my Ph.D., I am focusing on enhancing this therapy. My goal is twofold : first, to understand which clinical, physiological, and psychological factors affect this type of therapy, and second, to develop algorithms that are robust against these factors.

## Introduction

Brain-Computer Interfaces (BCI) are technologies that monitor (e.g., using electroencephalography), process (using machine learning techniques), and translate patterns of brain activity into commands for various digital technologies. They are utilized in post-stroke motor rehabilitation to deliver optimal time-matched visual and/or somatosensory feedback based on the activity in the somatosensory and motor cortices. The effectiveness of BCIs in enhancing plasticity and motor rehabilitation post-stroke has begun to emerge from extensive research in this field. The literature suggests that motor imagery or motor attempt training procedures significantly enhance motor function improvement compared to solely employing motor imagery or muscle and proprioception stimulation [1]. BCIs facilitate the co-activation of both top-down processes (i.e., motor imagination or attempts) and bottom-up processes (i.e., coherent somatosensory afferences resulting from visual or somatosensory stimulation of the affected limb). To advance current therapies, two critical aspects require attention. Firstly, understanding the factors that influence motor rehabilitation is essential for selecting the most appropriate therapy tailored to the patient's characteristics. Secondly, there is a need to refine the classification algorithms used in these therapies to enhance patient agency during the treatment process. During my Ph.D., I focused on those questions aimed at understanding and improving BCI-based motor rehabilitation after a stroke.

## Potential predictors that influence motor rehabilitation after a stroke.

Initially, we conducted an extensive literature review to identify factors influencing motor rehabilitation therapy. This included factors affecting the control of BCIs

for neurotypical users and factors known as predictor of motor rehabilitation for stroke patients without BCI-based therapy. The potential predictors of rehabilitation can be either psychological factors, mental state, neurophysiological predictors, demographical or medical information. This comprehensive synthesis helped us identify critical factors potentially influential in BCI-based therapy, which need to be validated in medical experiment.

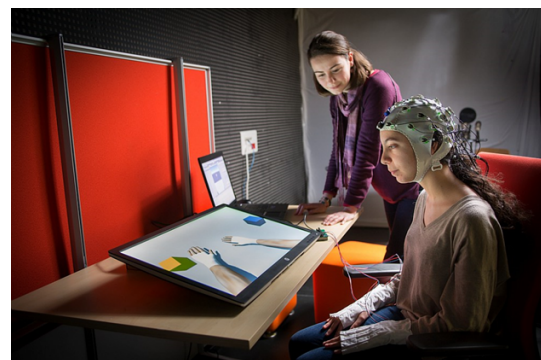


FIGURE 29.1 – Illustration of the experimental setup during MI-BCI based therapy. Left and right hand motor imagery is decoded with the electroencephalographie and is displayed on the screen.

## Developing new classification algorithms robust to factors of variability in BCI classification

Subsequently, we developed a modification of the Linear Discriminant Analysis algorithm (LDA), termed Invariant Linear Discriminant Analysis (iLDA) and presented in Algorithm 1. This model, robust against covariates known to influence BCI control, showed significant improvements in classification accuracies on simulated data. However, it did not demonstrate enhanced classification performance in EEG data considering the iden-

tified influential covariates.

---

**Algorithm 1 ILDA**


---

```

1: procedure FIT( $x, y, covariates$ )
2:   for  $classes = 1, 2$  do
3:     Compute  $A_{classes}$  and  $b_{classes}$  with the multivariate linear regression between  $X_{classes}$  and the corresponding  $Covariates$ 
4:   end for
5:    $A = (A_1 + A_2)/2$ 
6:    $b = (b_1 + b_2)/2$ 
7:   for  $classes = 1, 2$  do
8:      $\tilde{x}_{classes} = x_{classes} - (Covariates \cdot A' + b)$ 
9:   end for
10:  Compute  $\tilde{x}_i$  the variable sample where the influence of the  $covariates$  are linearly removed
11:  Compute  $\tilde{x}_{i_{classes}}$  the mean of the signal for on the time dimension
12:  Compute  $Cov_{classes}$  the covariance matrix of  $\tilde{x}_{i_{classes}}$ 
13:  end for
14:   $Cov = (Cov_1 + Cov_2)/2$ 
15:  Compute  $Inv_{cov}$  the inverse of the matrix  $Cov$ 
16:   $intercept = -(1/2) * (\tilde{x}_{i_1} + \tilde{x}_{i_2}) \cdot Inv_{cov} \cdot (\tilde{x}_{i_1} - \tilde{x}_{i_2})'$ 
17:   $weight = Inv_{cov} \cdot (\tilde{x}_{i_1} - \tilde{x}_{i_2})'$ 
18: end procedure

1: procedure PREDICT( $x, covariates$ )
2:    $linearprojection = intercept + weight \cdot (x - (covariates \cdot A' + b))'$ 
3:   if  $linearprojection < 0$  then
4:     predict class 1
5:   else
6:     predict class 2
7:   end if
8: end procedure

```

---

Exploring new axis for developing robust algorithms, I focused on deep learning techniques for EEG signal classification. We confirmed that deep learning techniques surpass linear models in classification accuracy on our open-source dataset. Interestingly, factors correlating with BCI performance in linear models were no longer significant using deep learning models, suggesting that the latter might extract more complex, non-linear features robust against known linear covariates. Deep learning models could thus enhance BCI control in stroke rehabilitation, offering resistance to factors traditionally influencing motor rehabilitation.

### **Validating efficacy of deep learning classifier in an online experiment**

We aim to compare the performance and user experience with two different classifiers, one linear classifier, Filter Bank Common Spatial Pattern (FBCSP) combine with (LDA), and a deep learning classifier, Deepconvnet [2]. Both model will be evaluated in the same MI-based BCI training with healthy participants. As deep learning models need a large amount of data and are able to generalize patterns across subjects, the model will be trained on a dataset from 87 participants performing the same motor imagery tasks (subject-independent design). FBCSP+LDA will be trained using calibration data from each subject, as they have poor generalization performance (subject-specific design). Our experiment will involve 20 subjects participating in six runs : two calibration runs with sham feedback, two runs utilizing the DeepConvNet classifier with feedback, and two runs with the FBCSP+LDA classifier with feedback, in a random order and unknown to the participants. We will evaluate the classifiers in terms of performance and user experience metrics including anxiety, agency, fatigue, frustration, subjective user performance, feedback com-

prehension, and feedback informativity.

### **Validating predictors of BCI-based motor rehabilitation after a stroke**

Lastly, we developed an experimental protocol scheduled to be conducted at CHU Pellegrin. This protocol involves BCI-based motor rehabilitation therapy for chronic-phase stroke patients. Addressing the question of factors influencing post-stroke motor rehabilitation. We planned to include 20 participants. Before, directly at the end and 3 months after the BCI therapy, the functional abilities (EmNSA and UE FMA) and the physiological profile (TMS measures (e.g. amplitude ; cortical excitability)) of the participants will be evaluated. Once included in the study, patients will perform 10 BCI training sessions based on motor attempt of around 1 hour each. Additionally, clinical, physiological and psychological profile of the patients will be evaluated through the completion of several questionnaires, tests and neurophysiological analysis. First, the psychological profile of the participants will be evaluated through the Mental Rotation Questionnaire, the 16PF5 questionnaire and the Kinaesthetic and Visual Imagery Questionnaire. Second, the neurophysiological and anatomical profile of the patients will be assessed through lesion characteristics on MRI, SSEP, EEG and MEP. This study aim to identify the potential predictors of motor rehabilitation performances by comparing the evolution of the motricity after the therapy and the different predictors that we extracted during the study. Those finding will allow us to propose recommendation to who are the patient that are more suitable to benefits from the therapy and create algorithms that take into account those characteristic into account and therefore create an adapted and personalized therapy for the patients.

### **Bibliographie**

- [1] Cervera, María A., et al. «Brain-Computer Interfaces for Post-Stroke Motor Rehabilitation : A Meta-Analysis». *Annals of Clinical and Translational Neurology*, vol. 5, n 5, mai 2018, p. 651-63.
- [2] Schirrmeister, Robin Tibor, et al. «Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization». *Human Brain Mapping*, vol. 38, n 11, novembre 2017, p. 5391-420. Pub-Med, <https://doi.org/10.1002/hbm.23730>.



# RENDU EXPRESSIF ET CONTRÔLE DE LIGNES

Panagiotis TSIAPKOLIS

Direction : Pierre BÉNARD

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

**L**es lignes sont un aspect important de beaucoup de styles de dessin, mais les intégrer dans un rendu n'est pas simple. Les détecter est un problème complexe, et contrôler leur apparence est encore un problème ouvert. Faire tout ça, en temps réel critique ( $<12\text{ms}$ ), est un tour de force, mais c'est les conditions pour pouvoir l'utiliser dans des applications comme le jeu vidéo. Cette thèse cherche à voir comment cela pourrait être possible.

## Rendu de Lignes

### Vue d'ensemble

Le rendu de ligne se passe en deux étapes conceptuelles principales :

- La détection, où l'algorithme doit trouver la position des lignes en utilisant l'image ou la scène 3D
- Le rendu, où les lignes sont choisies puis dessinées à l'écran selon plusieurs paramètres

Chaque étape a ses difficultés associées. La détection doit pouvoir trouver les lignes que l'artiste souhaite afficher, et donc doit pouvoir comprendre quelles lignes un artiste souhaite dessiner, tandis que le rendu doit pouvoir dessiner selon le style de l'artiste sans sa supervision.

### Types de lignes

Pour pouvoir comprendre mieux, Cole et al. [1] ont fait une étude en demandant à plusieurs artistes de dessiner par dessus un modèle 3D, et ont montré une forte corrélation entre des différences de gradients sur le modèle ou sur l'image et une ligne dessinée, ce qui permet de guider les algorithmes.

Afin de pouvoir les reproduire, les lignes peuvent être classés en plusieurs catégories (Fig. 30.1) en fonction du gradient sur lequel elles sont basées :

- Occlusion : quand un objet passe devant un autre, il s'agit d'un gradient de profondeur
- Courbature : quand une surface se plie, il s'agit d'un gradient de normales
- Séparation : quand un objet ou matériau change
- Illumination : présentes dans certains styles, il s'agit d'un gradient de lumière
- Arbitraires : lignes dessinées directement par l'artiste sans relation à l'objet, sont souvent utilisées pour représenter autre chose ou pour des explications

Ces lignes peuvent être associées aux gradients à plusieurs degrés de dérivation, et certaines sont dépendantes de paramètres externes comme du point de vue

ou de la lumière. Par exemple, beaucoup de lignes de courbature peuvent être précalculer, mais les lignes d'occlusion doivent être calculées à chaque rendu. Nous nous concentrons sur ces dernières.

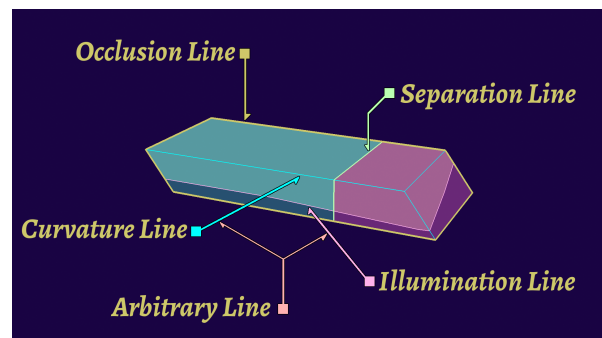


FIGURE 30.1 – Explication visuelle des types de lignes présents sur un rendu de gomme. Les lignes arbitraires ici sont les lignes utilisées pour expliquer les lignes d'autres types. Image par Panthavma [4]

## Détection

### Méthodes Existantes

Il existe trois méthodes principales pour la détection de contours d'occlusion sur modèles 3D. Parmi-elles, deux sont utilisables en temps réel.

**Inverted Hull.** Il s'agit d'une méthode rapide à mettre en place utilisant le backface culling du GPU. En créant une copie du maillage, l'agrandissant, et en inversant les normales, il est possible de faire une sorte de "coquille" autour du maillage qui s'affichera autour de lui. Cette méthode permet un bon contrôle de la taille et couleur des lignes, mais est limitée aux contours d'occlusion et présente de nombreux artéfacts. Des méthodes plus avancées utilisent les geometry shaders pour offrir plus de contrôle direct à l'utilisateur.

**Extraction en Espace Ecran.** Cette méthode consiste à utiliser des convolutions afin de trouver les discontinuités en espace écran, à travers des buffers spécialisés (profondeur, normales, object ID). Il s'agit d'une méthode très rapide et capable de trouver beaucoup de



types de lignes différents, mais elle est difficile à contrôler, et a du mal avec les lignes épaisses, demandant des heuristiques et des algorithmes spécialisés.

### Mesh Contours

La dernière méthode est les meshes contours [2], qui sont le sujet de ces travaux. Au lieu de rechercher les discontinuités en espace écran, on va tester chaque arête du maillage individuellement afin de déterminer les contours. Cette méthode permet une détection plus précise, ainsi que beaucoup plus d'options de style à travers ses primitives de ligne au lieu de pixels, mais est trop lente pour le temps réel.

**Algorithme.** Les meshes contours fonctionnent en 4 étapes :

- Détection : On vérifie chaque arête en regardant si une de ses faces regarde la caméra, et l'autre non. Si c'est le cas, l'arête est considérée comme contour et passe à l'étape suivante. (Fig. 30.2)
- Visibilité : Les contours détectés ainsi peuvent être cachés derrière l'objet. Une étape est nécessaire pour filtrer les lignes visibles et invisibles.
- Chaînage : Les segments de ligne sont encore séparés à ce moment, donc pour pouvoir faire des stylisations plus avancées, il est nécessaire de les assembler en des lignes plus grandes.
- Rendu : Une fois ces lignes assemblées, on peut désormais les afficher selon différents paramètres de style. L'utilisation de primitive de ligne permet une meilleure qualité que les méthodes précédentes.

La majorité de nos travaux repose sur la première étape de détection, étant lente par elle même.

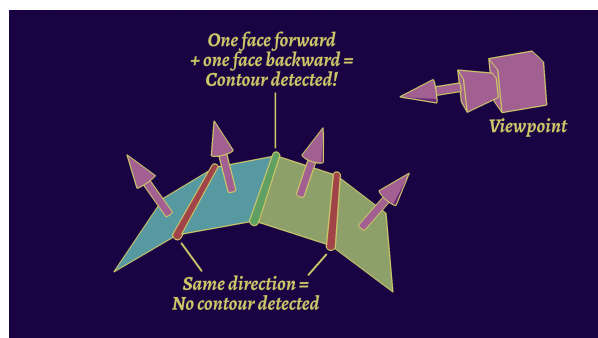


FIGURE 30.2 – Explication visuelle du test de contour. Une arête est considérée comme contour uniquement dans la condition où une face regarde la caméra et l'autre non. Image par Panthavma [4]

**Accélération.** Afin de pouvoir réaliser le calcul dans un temps raisonnable, il est nécessaire d'éliminer une partie des tests avant leur réalisation. Les structures d'accélération existantes étaient faites pour le CPU, avec des structures hiérarchiques non-adaptés au GPU [5], et limités aux modèles rigides.

**GPU.** Quelques travaux récents ont appliqué les meshes contours sur GPU en temps raisonnable [3], mais l'étape d'extraction était basique et limitée aux modèles rigides.

**Notre méthode.** Nous avons amélioré ces méthodes en proposant une structure d'accélération adaptée au GPU, avec des tests de performance plus poussés. Notre contribution se concentre sur la méthode de construction de cette structure, qui utilise de nouvelles avancées mathématiques pour permettre un groupement plus efficace, dont les résultats sont dans la table 30.1. Ces travaux sont en cours de publication.

| Méthode             | Temps     |
|---------------------|-----------|
| Bruteforce (CPU)    | 1259.13ms |
| SNCH (CPU) [5]      | 808.76ms  |
| Notre méthode (CPU) | 290.66ms  |
| Bruteforce (GPU)    | 88.50µs   |
| Notre méthode (GPU) | 29.70µs   |

TABLE 30.1 – Performance comparison sur le modèle du Stanford Armadillo

### Bibliographie

- [1] F. Cole et al., “Where Do People Draw Lines?,” p. 11, 2008.
- [2] P. Bénard and A. Hertzmann, “Line Drawings from 3D Models : A Tutorial,” FNT in Computer Graphics and Vision, vol. 11, no. 1–2, pp. 1–159, 2019, doi : 10.1561/06000000075.
- [3] W. Jiang, G. Li, Y. Nie, and C. Xian, “GPU-Driven Real-Time Mesh Contour Vectorization,” p. 13, 2022.
- [4] Panthavma. “Line Rendering Deep Overview Part 1 - Extraction” <https://panthavma.com/articles/lines/deep-overview-extraction>
- [5] D. E. Johnson and E. Cohen, “Spatialized Normal Cone Hierarchies,” 2001.

# EEG-BASED DECODING OF AESTHETIC PREFERENCE FOR VISUAL ART

Marc WELTER

Direction : Fabien LOTTE

Univ. Bordeaux, CNRS, Bordeaux INP, INRIA, LaBRI, UMR 5800, F-33400 Talence, France

**P**assive Brain-Computer-Interfaces (BCIs) able to decode aesthetic preference could improve user experience in digital spaces by personalizing aesthetic stimuli selection without requiring explicit user feedback that might interrupt aesthetic experience. However, neuroscientific understanding of aesthetic experience remains lacking, while the tried and tested BCI classification algorithms have not yet been applied to decode aesthetic preferences from EEG signals. We thus conducted an experiment in which participants were exposed to visual artworks in a virtual museum and requested to grade their preferences for each of them, all this while their EEG was being measured. Previous neuroaesthetic research suggested that oscillatory modulations in different neural frequency bands could be informative of aesthetic preference. Therefore, we tested a time-frequency feature classification method widely used in BCIs, i.e. Filterbank Common Spatial Patterns feature extraction together with shrinkage Linear Discriminant Analysis, in a 2-class aesthetic preference classification problem. We report promising aesthetic preference decoding accuracies significantly and substantially above chance level.

## Introduction

Humans today are surrounded by a vast amount of artworks in digital spaces, e.g. social media. In such spaces, the presentation of art, e.g. visual art or music, often relies on preference-based recommendation systems that require explicit feedback from the user. However, explicit feedback, e.g. clicking on a Like button, requires cognitive effort which can diminish the aesthetic experience.

Brain-Computer-Interfaces (BCIs) able to decode aesthetic preference could improve user experience in digital spaces by personalizing aesthetic stimuli selection without requiring explicit user feedback [1]. However, neuroscientific understanding of aesthetic experience remains lacking, while the tried and tested BCI classification algorithms have not yet been applied to decode aesthetic preferences from Electroencephalography (EEG) signals [2]. Therefore, we have been investigating if commonly used BCI methods can be applied in aesthetic preference decoding.

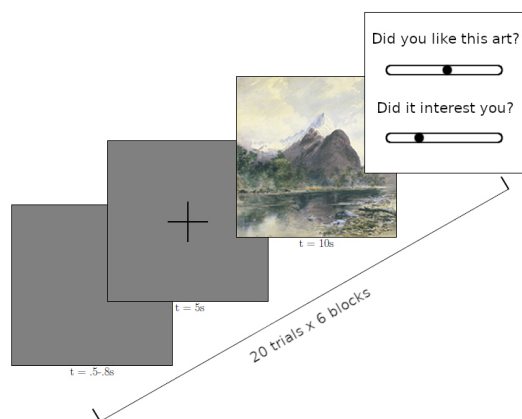
## Materials and Methods

We established a diverse dataset of more than 200 high quality digital reproductions of artworks originating from almost all continents and ranging from pre-historic to contemporary time periods in the public domain. Figure 31.1 shows one participant gazing at AI-generated art during data recording. We conducted an experiment in which 13 participants were exposed to 120 visual artworks in a virtual museum. The participants were asked to grade their preferences for each of them, all this while their EEG was being measured. Figure 31.2 shows the experimental protocol.



FIGURE 31.1 – A participant taking part in the aesthetic preference decoding BCI experiment

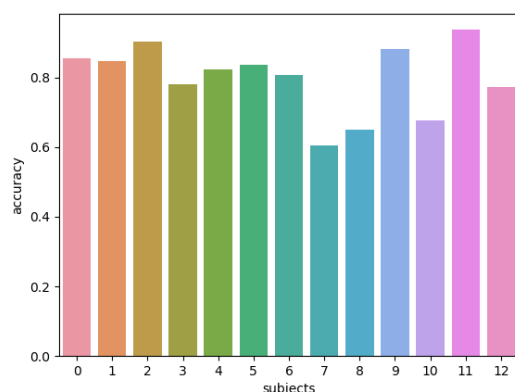
Previous neuroaesthetic research suggested that oscillatory modulations in different neural frequency bands could be informative of aesthetic preference. Therefore, we tested a time-frequency feature classification method widely used in BCIs, i.e. Filterbank Common Spatial Patterns (FBCSP) [3] feature extraction together with shrinkage Linear Discriminant Analysis (sLDA) [4], in a 2-class (Like vs Dislike) single-trial aesthetic preference classification problem. EEG Artefacts were removed using Independent Component Analysis prior to feature extraction. Chance level classification performance for 2-class classification with a minimum of 55 samples per class was estimated at 60.9%. Classification performance was validated with a 10-fold shuffled and stratified cross-validation. In addition, we also ran a 10-fold shuffled and stratified cross-validated CSP-sLDA classification for each band-pass filter individually, in order to investigate the discriminatory power of each frequency band.



**FIGURE 31.2** – The data collection included 6 blocks of 20 visual art stimuli presentations and subsequent subjective aesthetic evaluations

## Results

For FBCSP-sLDA, we report promising offline aesthetic preference decoding accuracies above chance level with a mean of  $0.798\% \pm 0.162$ . Figure 31.3 shows the mean FBCSP-sLDA classification performance for all subjects



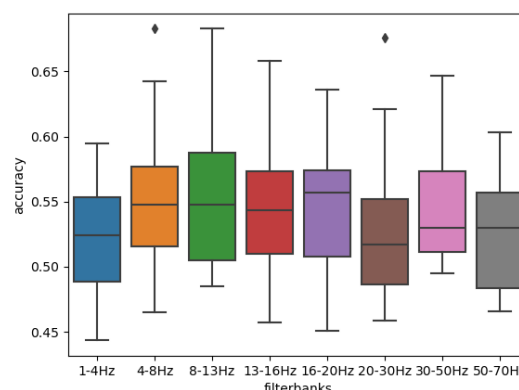
**FIGURE 31.3** – Mean FBCSP-sLDA Aesthetic Preference Classification Scores For All Subjects

CSP-sLDA classification on individual EEG frequency bands performed much worse and well below the estimated chance level on average with a mean accuracy of  $0.541\% \pm 0.052$  and showed a large variability across subjects.

## Discussion

Despite the relatively high FBCSP-sLDA classification performance, closing the gap towards online aesthetic preference decoding, and thus automated art presentation BCIs, remains challenging. Many BCIs require training on balanced data samples. However, it is often uncertain beforehand whether an individual will like an artwork or not. Objective properties of an artwork, such as symmetry or complexity, have been correlated with subjective aesthetic ratings [5] and could be used to estimate aesthetic preference a priori. In conclusion, aesthetic preference decoding BCIs seem feasible, but

future work needs to investigate optimal training data selection during online calibration.



**FIGURE 31.4** – Variability of CSP-sLDA classification performance between subjects for all frequency bands

## Bibliography

- [1] Welter M., Bouneau A. Lotte F. & Ward T. (2023). Towards curating personalized art exhibitions in Virtual Reality with multimodal Electroencephalography-based Brain-Computer-Interfaces. *Art & Perception*, 11 (2). doi : <https://doi.org/10.1163/22134913-20230001>
- [2] Welter M. & Lotte F. (2024). Ecological decoding of visual aesthetic preference with oscillatory electroencephalogram features—a mini-review. *Frontiers in Neuroergonomics*, 5. doi : <https://doi.org/10.3389/fnrgo.2024.1341790>
- [3] Ang, K., Chin, Z., Wang, C., Guan, C. & Zhang, H. (2012). Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b. *Frontiers in neuroscience*, 6 (59). doi : 39. 10.3389/fnins.2012.00039.
- [4] Blankertz, B., Lemm, S, Treder, M., Haufe, S & Müller, KR. (2010). Single-trial analysis and classification of ERP components — A tutorial. *neuroimage*, 56. doi : 10.1016/j.neuroimage.2010.06.048
- [5] Geller HA., Bartho R., Thoemmes K. & Redies C.(2022). Statistical image properties predict aesthetic ratings in abstract paintings created by neural style transfer. *Frontiers in Neuroscience*, 16. doi : 10.3389/fnins.2022.999720

# CYLINDERS IN SQUARE-TILED SURFACES AND METRIC RIBBON GRAPHS

Ivan YAKOVLEV

Direction : Vincent DELECROIX

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

**M**y thesis presents several contributions to the refined asymptotic enumeration of square-tiled surfaces. Using a natural geometric decomposition, this problem reduces to the study of counting functions for metric ribbon graphs. We use purely combinatorial tools to compute some of these counting functions. This allows us to generalize formulas previously obtained by algebraic methods.

## Introduction

**Square-tiled surfaces.** A square-tiled (ST) surface is a closed oriented surface constructed from a finite collection of Euclidean unit squares  $[0, 1]^2$  by isometrically identifying their sides and respecting the following gluing rule : sides are only identified in pairs of type top/bottom or left/right (see Figure 32.1 for an example). A ST surface is equipped with a flat metric (coming from the squares) except at the possible conical singularities – points where more than 4 vertices of squares are identified. The gluing rule implies that the holonomy of this singular flat metric is trivial.

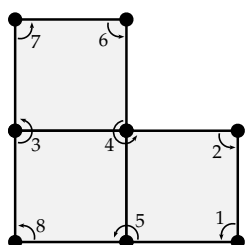


FIGURE 32.1 – Identifying the opposite sides of this polygon, one obtains a surface of genus 2 tiled with 3 squares. All of the vertices of the squares are identified together into one conical singularity of angle  $12 \cdot \frac{\pi}{2} = 6\pi$ . The numbered arrows represent the order in which the corners are visited when going around the singularity.

**Masur–Veech volumes.** The motivation for the study of ST surfaces comes from the theory of translation surfaces – arbitrary flat surfaces with conical singularities and trivial holonomy. They come in continuous families called strata. It turns out that asymptotic enumeration of ST surfaces in a given stratum is equivalent to the computation of its volume (Masur–Veech volumes). Precise knowledge of these volumes is important for the study of a natural dynamical system on the strata (Teichmüller flow), with applications to billiards in rational polygons.

**Algebraic vs combinatorial approach.** Despite the combinatorial nature of the problem, explicit generating functions and recursions for the volumes were obtained

via algebraic geometry [1, 2] by identifying them with certain intersection numbers on the strata. Nevertheless, a combinatorial approach was recently revived [3] in the context of half-translation surfaces (here the holonomy of the surface is allowed to be  $\mathbb{Z}/2\mathbb{Z}$ ), motivation being the high genus asymptotic analysis of the volumes.

**Cylinder decomposition of ST surfaces.** The general strategy in [3] is to decompose (in a canonical way) a ST surface into simple geometric pieces (cylinders) glued along graphs embedded into the surface (metric ribbon graphs, see Figure 32.2). The enumeration of ST surfaces then reduces to the study of the counting functions for metric ribbon graphs. The volumes are decomposed into a sum of contributions of surfaces of different combinatorial types. When divided by the total volume, these contributions can be interpreted as the probabilities of a random ST surface to have a given type.

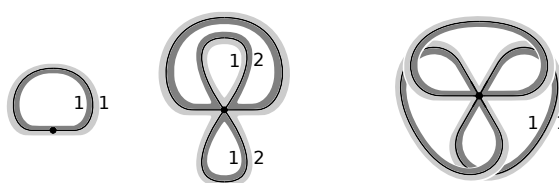


FIGURE 32.2 – Examples of ribbon graphs with one vertex and labeled boundary components of two colors.

## Main result

**Cylinders in ST surfaces of minimal strata.** The idea of my paper [5] is to use the strategy of [3] in the context of translation surfaces. The motivation is to prove the following conjecture made by the same authors.

**Conjecture 1.** *The distribution of the number of cylinders in a random ST surface of genus  $g$  is asymptotically equivalent when  $g \rightarrow \infty$  (in a certain strong sense) to the Poisson distribution with parameter  $\log(4g - 3)$ . This approximation is uniform over all (non-hyperelliptic) connected components of all strata.*



The natural place to start are the minimal strata, where surfaces have exactly one singularity. The main result of my paper [5] is the following exact enumeration formula.

**Théorème 5 ([5]).** *The contribution of  $n$ -cylinder  $ST$  surfaces to the volume of the minimal stratum in genus  $g$  is equal to  $\frac{2(2\pi)^{2g}}{(2g-1)!}a_{g,n}$ , where the numbers  $a_{g,n} \in \mathbb{Q}$ , and whose generating function  $\mathcal{C}(t, u) = 1 + \sum_{g \geq 1} (\sum_{n=1}^g a_{g,n} u^n) (2g-1)t^{2g}$  satisfies for all  $g \geq 0$  :  $\frac{1}{(2g)!} [t^{2g}] \mathcal{C}(t, u)^{2g} = [t^{2g}] \left( \frac{t/2}{\sin(t/2)} \right)^u$ . Here  $[t^{2g}]$  stands for the extraction of the corresponding coefficient.*

Setting  $u = 1$  in Theorem 5, we recover the generating function for the total volumes of the minimal strata, obtained previously by Sauvaget [1] via intersection theory.

**A word about the proof.** In order to obtain this result, I had to study the counting functions for a certain family of metric ribbon graphs which has not been previously considered. The main tool for this study was the bijection of Chapuy, Féray and Fusy [4] between ribbon graphs with one boundary and properly decorated plane trees, which allows for explicit enumeration of the former. The main complication was the presence of metric on the ribbon graphs, which must also satisfy certain linear conditions on the lengths of the edges. It turned out that the bijection also allows to control the metric. This led to an unexpected polynomiality of the top-degree term of the counting function, which is a priori only piecewise polynomial (this phenomenon also occurs for the counting functions in [3], the Kontsevich polynomials).

## Further results

**Cylinders in connected components of minimal strata.** Generically, each minimal stratum of translation surfaces has 3 connected components. One of them, the hyperelliptic component, represents surfaces with a symmetry of order 2, and is easy to analyze. The other two are distinguished by a topological invariant, called the parity of the spin structure.

It turns out that the approach of my paper [5] can be generalized to find the cylinder statistics as in Theorem 5, but for each of the connected components. This boils down to finding a combinatorial analog of the aforementioned topological invariant for ribbon graphs and computing the corresponding counting functions, whose top-degree terms also turn out to be polynomial. The obtained explicit formula also gives a refinement of a formula for the total volumes previously obtained by Chen, Möller, Sauvaget and Zagier [2] algebraically.

**Volumes of odd strata of half-translation surfaces.** This is joint work with Eduard Duryev and Élise Goujard. The aim here is to generalize the combinatorial formulas of [3] for the volumes of principal strata of half-translation surfaces to more general strata (with the degrees of all singularities being odd).

The motivation is that the necessary counting functions for metric ribbon graphs are known, and are the generalizations of the ones from [3]. What makes the case of non-principal strata more difficult, is that the direct application of the previous strategy actually overcounts ribbon graphs – one also obtains certain degenerate, disconnected ones.

My contribution to the project is a combinatorial proof of the formula for these degeneration coefficients. It is a generalization of similar computations performed in my paper [5]. Here, the degenerate ribbon graphs form tree-like structures but, again, because of certain linear constraints on the metric, not all trees appear. I succeeded in proving the formula by devising an analog of the well-known Prüfer code for trees, which takes into account the constraints on the metric.

**Triangulations of the product of simplices.** While studying the spin parity invariant for ribbon graphs, I found a family of (asymptotically)  $(n+m-1)!$  different triangulations of the product of simplices  $\Delta^n \times \Delta^m$ , which come from certain collections of labeled plane trees. What is surprising here is that simplices on the vertices of  $\Delta^n \times \Delta^m$  correspond naturally to non-plane trees. The additional planar structure produces a family of triangulations which seems to have not been considered before.

The interest here comes from the fact that, despite the simplicity of the polytope, the set of its triangulations is not completely understood : only a couple of explicit constructions are known ; describing the structure of the secondary polytope of  $\Delta^n \times \Delta^m$  is a long-standing open problem.

## Bibliographie

- [1] Adrien Sauvaget. Volumes and Siegel-Veech constants of  $\mathcal{H}(2g-2)$  and Hodge integrals. *Geom. Funct. Anal.*, 28(6) :1756–1779, 2018.
- [2] Dawei Chen, Martin Möller, Adrien Sauvaget, and Don Zagier. Masur-Veech volumes and intersection theory on moduli spaces of abelian differentials. *Invent. Math.*, 222(1) :283–373, 2020.
- [3] Vincent Delecroix, Élise Goujard, Peter Zograf, and Anton Zorich. Masur-Veech volumes, frequencies of simple closed geodesics, and intersection numbers of moduli spaces of curves. *Duke Math. J.*, 170(12) :2633–2718, 2021.
- [4] Guillaume Chapuy, Valentin Féray, and Éric Fusy. A simple model of trees for unicellular maps. *J. Comb. Theory, Ser. A*, 120(8) :2064–2092, 2013.
- [5] Ivan Yakovlev. Contribution of  $n$ -cylinder square-tiled surfaces to Masur-Veech volume of  $\mathcal{H}(2g-2)$ . *Geom. Funct. Anal.* 33, 1682-1716, 2023.



# ROCKMATE : MEMORY EFFICIENT TRAINING FOR NEURAL NETWORKS

Xunyi ZHAO

Direction : Olivier BEAUMONT  
Encadrement : Lionel EYRAUD-DUBOIS

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

**W**e present ROCKMATE, an efficient memory-reducing framework to enable fine-tuning large language models on a single GPU. ROCKMATE uses integer linear programming to combine re-materialization, activation offloading, and CPU optimization in a holistically optimized way, ensuring an efficient usage of available resources. In the same way that PyTorch Dynamo takes a model and automatically changes it to reduce the execution time, ROCKMATE takes a model and automatically adapts it to fit memory constraints. This makes it possible to perform a full model fine-tuning of Llama-7B in full precision on a 12GB GPU, which originally requires 100GB of memory, with very limited communication overhead. ROCKMATE is designed to be compatible with reduced precision and parameter-efficient fine-tuning techniques, so that the memory benefits can be combined.

## Introduction

Large Language Models (LLMs) have demonstrated good performance in various language-related domains, including question answering, code generation, and natural language translation. Inspired by the transformer structure, LLMs with different architectures such as GPT [2], Bloom [1], Llama [3] and Mistral [4] have been trained and demonstrate excellent performance on general tasks. Based on these large pre-trained models, which require a significant amount of resources to train, many fine-tuned models have been proposed and contribute to a variety of fields. Compared to training an LLM from scratch, fine-tuning is more accessible to broader communities of artificial intelligence (AI) enthusiasts because it requires less data, less computing power, and less training time.

However, a common challenge for fine-tuning LLMs is the memory bottleneck of the training process. Most billion-parameter models can hardly be stored on consumer graphics cards, which typically have between 8GB and 24GB of video RAM (VRAM). Depending on the choice of hyperparameters and optimization settings, the training process can require far more memory than the VRAM available on a single GPU. While most LLMs are pre-trained using parallel approaches across hundreds of GPUs, many works [6, 7] have been proposed to allow fine-tuning of LLMs on a single GPU. Most methods reduce memory requirements by simplifying the training process, such as lowering the data precision or reducing the number of trainable parameters. These strategies have proven to be efficient in many cases without degrading the accuracy of the resulting model.

## Strategies

The memory footprint when training a large model consists of two parts : (1) intermediate activations  $M_{act}$ , whose size depends on the input batch size and which can be efficiently rebuilt using re-materialization to reduce the associated peak memory usage ; (2) model states, which consist of parameters  $M_{param}$ , parameter gradients  $M_{p\_grad}$ , and optimizer states  $M_{opt\_st}$ . Once the model is selected, the size of  $M_{param}$  depends only on the data type of the parameters. The size of  $M_{p\_grad}$  is the size of trainable parameters, which can be much smaller than  $M_{param}$  when using PEFT methods. The size of  $M_{opt\_st}$  depends on the choice of the optimizer and is typically a multiplicative factor of  $M_{p\_grad}$  : for the Adam family of optimizers, the multiplicative factor is 2. For example, the relative sizes of each part when training a full-precision Llama2-7B model with batch size 2 and sequence length 512 are  $M_{param} = M_{p\_grad} = 25\text{GB}$  and  $M_{act} = 24\text{GB}$ . If the model is optimized with the Adam optimizer [5],  $M_{opt\_st} = 50\text{GB}$ , so that the total memory requirement is 124GB. If gradient accumulation is not used, parameter gradients can be removed from memory at each iteration,  $M_{p\_grad}$  and  $M_{act}$  are not always required simultaneously. Still, the peak memory usage for such a training task is expected to be 100GB, which is not feasible on any consumer GPU.

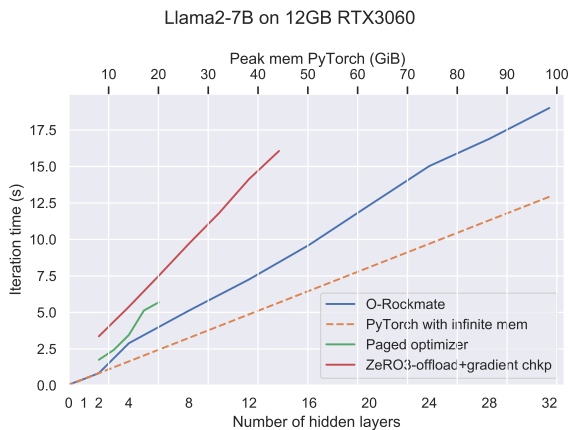
In ROCKMATE, we adopt the same assumption of a block sequential structure that is relevant for many Large Language Models, but we aim to reduce all sources of memory consumption. To reduce  $M_{act}$ , we use the RK-CHECKMATE formulation to generate multiple re-materialization schedules for each block. Instead of applying ROCKMATE Phase (2), we combine these schedules with a new linear programming formulation called

O-ILP, which incorporates techniques for reducing all other sources of memory consumption.

To reduce  $M_{param}$ , each parameter can be offloaded onto CPU RAM. To reduce  $M_{p\_grad}$ , each parameter is optimized immediately when the associated gradients are generated and no gradient accumulation is applied. To reduce  $M_{opt\_st}$ , since the basic variant of O-ILP performs all optimization steps onto GPU, the optimizer states can either be kept on GPU or offloaded to CPU RAM and moved back to GPU before the next optimization step (as proposed by Paged Optimizers [6]).

We also propose a "CPU optimization" variant of ROCKMATE that allows parameters to be optimized on CPU, in which case optimizer states will always be kept in CPU RAM. This idea was originally proposed in ZeRO-Infinity [7], where it was noted that since it is possible to (i) offload the gradient of a parameter, (ii) optimize on the CPU, and (iii) prefetch the new optimized parameter, CPU optimization involves the same amount of data transfers as regular offloading.

The main contribution of ROCKMATE is to perform a holistic optimization of all these techniques simultaneously, so that different parameters can be handled by different techniques to minimize the iteration time within a limited memory budget. This combination makes it possible to train the entire network within the memory requirements of a single block : in the context of very tight memory budgets, all data related to other blocks can indeed be offloaded to the CPU. When more memory is available, ROCKMATE automatically reduces the amount of offloaded data.



**FIGURE 33.1** – Average iteration time vs. number of hidden layers (Llama2-7B corresponds to 32 layers). Batch size is 4 and sequence length is 512, resulting in  $M_{act} = 45GB$ . Peak mem PyTorch represents the expected memory usage if GPU memory is infinite.

## Conclusion

We present ROCKMATE, a framework for efficiently reducing memory requirements when fine-tuning Large Language Models on a single consumer-grade GPU. ROCKMATE relies on integer linear programming to combine different memory-reducing approaches in a holistically optimized manner, ensuring efficient use of all available

resources. Experiments on typical fine-tuning scenarios show that ROCKMATE can achieve a  $10\times$  reduction in memory at the price of a  $1.5\times$  increase in training time, without modifying the training task (e.g., lowering data precision or reducing the number of trainable parameters). We also show that ROCKMATE can be combined with parameter-efficient fine-tuning methods such as LoRA. ROCKMATE is an easy-to-use framework that can enable fine-tuning on resource-constrained machines and is expected to have a significant impact for individual AI researchers.

Future research efforts could add support for gradient accumulation, allow fine-grained offloading within a block to reduce the minimum memory requirement, and include activation offloading to further improve resource utilization.

## Bibliographie

- [1] Workshop, B., Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A., Yvon, F. & Others Bloom : A 176b-parameter open-access multilingual language model. *ArXiv Preprint ArXiv :2211.05100*. (2022)
- [2] OpenAI, R. Gpt-4 technical report. arxiv 2303.08774. *View In Article*. 2 pp. 13 (2023)
- [3] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S. & Others Llama 2 : Open foundation and fine-tuned chat models. *ArXiv Preprint ArXiv :2307.09288*. (2023)
- [4] Jiang, A., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D., Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L. & Others Mistral 7B. *ArXiv Preprint ArXiv :2310.06825*. (2023)
- [5] Kingma, D. & Ba, J. Adam : A method for stochastic optimization. *ArXiv Preprint ArXiv :1412.6980*. (2014)
- [6] Dettmers, T., Pagnoni, A., Holtzman, A. & Zettlemoyer, L. Qlora : Efficient finetuning of quantized llms. *ArXiv Preprint ArXiv :2305.14314*. (2023)
- [7] Ren, J., Rajbhandari, S., Aminabadi, R., Ruwase, O., Yang, S., Zhang, M., Li, D. & He, Y. ZeRO-Offload : Democratizing Billion-Scale model training. *2021 USENIX Annual Technical Conference (USENIX ATC 21)*. pp. 551-564 (2021)

**Deuxième partie**

**Mathématiques**



# MODELING AND DECOMPOSITION METHODS FOR ROBUST FACILITY LOCATION AND NETWORK DESIGN PROBLEMS WITH INTEGER RECOURSE

Komlanvi Parfait AMETANA

Direction : Boris DETIENNE

Co-Direction : Olga BATTALA

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

In this thesis, we focus on facility location and network design problems that arise in disaster or relief management. To account for the uncertainty inherent in these problems, we use risk measures to assess the resilience of solutions. Given the critical and humanitarian importance of the problems considered, this thesis focuses on the use of robust optimization methods as a modeling and solution approach.

## Scientific Context

The optimization problems that can be modeled with a mixed integer linear program have been intensively studied in the last decades. Historically, because of the difficulty of these problems, mostly deterministic versions of these problems have been investigated : all data are supposed to be perfectly known at the moment of the decision. However, these hypotheses do not correspond to the reality of decision-making, where data may vary between the moment the decision is taken and the moment the decision is implemented.

To take into account the uncertainty of the input data, two main approaches are used in the literature : stochastic and robust optimization. Stochastic optimization requires knowledge about the probability distribution of the data, represented by stochastic variables. This approach aims at providing good solutions on average. Robust optimization (see e.g. [2]) provides solutions that will not be very bad, even in the worst-case scenario. This approach can be used even when the amount of data is not sufficient to infer relevant probability distributions for the input parameters.

An important part of the optimization problems under uncertainty can be modeled as two-stage or multi-stage optimization problems. At the first stage, or during planning, decisions must be taken based on uncertain data. At the second stage, or recourse, the real values of uncertain parameters are revealed and the decision is either updated or completed. In the case of a two-stage setting, the considered robust optimization problem takes the following *min-max-min* form :

$$\min_{x \in X} \max_{\xi \in \Xi} \min_{y \in Y(x, \xi)} \bar{f}(x, \xi, y)$$

Under this form, function  $\bar{f}$  integrates the necessary recourse to make the final solution feasible :  $\bar{f}(x, \xi, y) = f(x, \xi, y) + I_{X(\xi)}(x, y)$  where  $I_{X(\xi)}$  takes infinite value if its argument is not feasible under the realiza-

tion of  $\xi$ . Efficient solution methods have been developed for the special case of static robust problems  $\min_{x \in X} \max_{\xi \in \Xi} f(x) + I_{X(\xi)}(x)$ , where  $\Xi$  is a convex compact set and  $\min_{x \in X} f(x)$  is a mixed integer linear program (MILP) or a convex program [3]. However, the optimization problems with recourse (a fortiori with integer recourse) have been less frequently considered. To give an idea of the difficulty of such problems, let us mention that many deterministic problems belonging to class P have an NP-hard robust counterpart, and the problems considered here are already NP-hard in their deterministic version. Recent approaches aim at leveraging advanced MIP reformulation techniques to address these problems (see e.g. [5, 1]).

## Objectives

The main objective is to develop efficient solution methods for robust optimization problems with integer recourse. In particular, the problems of facility location and network design will be considered. The uncertainty setting will be inspired by disaster or relief management. Theoretical analysis of the complexity of such optimization problems will also be investigated.

## Contributions

In our work, we study the strategic design of blood supply-chain logistics to face disasters caused by earthquakes. The goal of this supply chain is to convey blood from a set of donors to a set of patients in hospitals. We develop two-stage and multi-period risk aversion models for the blood supply chain in disaster management, focusing on measures such as mean-CVaR and robust worst-case. The objective is to provide decision-makers with a comprehensive comparison of risk-averse models for the design of a blood supply chain to guide their choices for ethically acceptable optimization in the context of mitigating the damages caused by an earthquake. We therefore review the relevant existing literature, introduce the considered optimization problem,



present different mathematical models for risk aversion, and evaluate these models on an out-of-sample set. The numerical results indicate that our approaches manage risk more effectively than the solutions presented in the previous model [4].

Additionally, We provide insights for policymakers, healthcare practitioners, and logistics professionals to improve blood supply chain management strategies in disaster response and preparedness efforts.

In addition, we presented a simulation algorithm for generating a discrete set of scenarios to model uncertainty. This algorithm evaluates the impact of earthquake and aftershock scenarios on a population divided into groups over a given area. And by grouping, we obtain the demand for blood in hospitals.

The size of the uncertainty set affects the complexity of the models. Numerical results show that a coarse approximation of the uncertainty is more effective than a fine approximation for the problem at hand.

## Perspectives

Having demonstrated the effectiveness of the robust worst-case model, our next step will be to extend our study to solve a larger number of instances. We'll start by finding an efficient way to transition from a discrete to a polyhedral uncertainty set. Next, we will investigate the extent to which we can apply reformulation techniques such as K-adaptability, decision rules, and dualization. In parallel, we will try to improve the quality of the solution by using decomposition methods. These perspectives will allow us to extend our understanding and optimize the proposed solutions.

## Bibliographie

- [1] Arslan, A. N. and Detienne, B. (2022). Decomposition-based approaches for a class of two-stage robust binary optimization problems. *INFORMS Journal on Computing*, 34(2).
- [2] Ben-Tal, A., El Ghaoui, L., and Nemirovski A. S. Robust optimization. Princeton series in applied mathematics. Princeton University Press, Princeton, 2009. OCLC : ocn318672208.
- [3] Bertsimas, D. and Sim, M. (2004). The Price of Robustness : Operations Research. *Operations Research*, 52(1) :35-53.
- [4] Jabbarzadeh, A., Fahimnia, B., and Seuring, S. (2014). Dynamic supply chain network design for the supply of blood in disasters : A robust model with real world application. *Transportation Research Part E : Logistics and Transportation Review*, 70 :225-244
- [5] Subramanyam, A., Gounaris, C. E., and Wiesemann, W. (2018). K-Adaptability in Two-Stage Mixed-Integer Robust Optimization.

# CONDENSED DUALITY THEOREMS FOR THE WEIL GROUP OF A $p$ -ADIC FIELD

Marco ARTUSA

Direction : Baptiste MORIN

Univ. Bordeaux, Institut de Mathématiques de Bordeaux, UMR 521, F-33400 Talence, France

**D**uality is a mathematical principle allowing to give two points of view on the same object. Equivalently, information about a mathematical object can be obtained thanks to its "dual". This principle is present in different areas of mathematics (geometry, algebra, analysis). An example from geometry is the symmetrical relationship between points and lines in the projective plane : two different points can be joined by a unique line, and two different lines intersect in a unique point. Another example from algebra is the relationship between a finite-dimensional vector space and the vector space of linear forms on it. In the following we present an instance of duality theorem in number theory : the local Tate duality. Moreover, we see how this theorem can be enlarged.

## Some expressions of duality

We use the example of finite-dimensional vector spaces to show that some mathematical dualities can be expressed both as perfect pairings and as equivalence of categories.

Let  $M, N$  and  $D$  be abelian groups. A pairing is a biadditive map  $M \times N \rightarrow D$ . This pairing is said to be perfect if the induced morphisms  $M \rightarrow \text{Hom}(N, D)$  and  $N \rightarrow \text{Hom}(M, D)$  are isomorphisms. In this case,  $N$  is the dual of  $M$  and viceversa, and  $D$  is the dualising abelian group. We can replace abelian groups by vector spaces (or objects of other categories) in the previous discussion. If  $k$  is a field and  $V$  is a finite-dimensional  $k$ -vector space, the valuation pairing

$$V \times \text{Hom}_k(V, k) \rightarrow k, \quad (v, f) \mapsto f(v).$$

expresses the duality between  $V$  and  $\text{Hom}_k(V, k)$ .

We can also express this duality in a categorical way as follows. If we denote by  $\text{Vect}_k^{fd}$  the category of finite-dimensional  $k$ -vector spaces, the functor

$$(-)^* : (\text{Vect}_k^{fd})^{op} \rightarrow \text{Vect}_k^{fd}, \quad V \mapsto V^* := \text{Hom}_k(V, k)$$

is an equivalence of categories. This means that there is an isomorphism between  $V$  and  $(V^*)^*$ , natural in  $V$ .

### The Pontryagin duality

Locally compact abelian groups are topological abelian groups whose underlying space is locally compact. For any locally compact abelian group  $A$ , we endow the abelian group  $\text{Hom}(A, \mathbb{R}/\mathbb{Z})$  of continuous homomorphisms  $A \rightarrow \mathbb{R}/\mathbb{Z}$  with the compact-open topology. We obtain a topological abelian group  $A^\vee$ , the Pontryagin dual.

Pontryagin duality states that  $A^\vee$  is a locally compact abelian group if  $A$  is. Moreover, the functor  $(-)^\vee : \text{LCA}^{op} \rightarrow \text{LCA}$  is an equivalence of categories. This restricts to a duality between discrete abelian groups and compact abelian groups.

## A Duality Theorem in Arithmetics

In number theory, important duality theorems due to Tate and Poitou are collected in [4]. They deal with Galois cohomology, an arithmetic invariant associated to a field  $F$ .

### Galois cohomology

Let  $F$  be a field and let  $\bar{F}$  be its algebraic closure. The Galois group  $G_F := \text{Aut}_F(\bar{F})$  is the group of automorphisms of  $\bar{F}$  which fix  $F$ . Topologically, it is a profinite group.

This group acts continuously on some sets coming from arithmetics. For example, if  $Z_f(\bar{F}) \subset \bar{F}^n$  is the set of solutions in  $\bar{F}^n$  of a polynomial  $f \in F[t_1, \dots, t_n]$ , then  $G_F$  acts on  $Z_f(\bar{F})$ , and the set of fixed points is  $Z_f(F)$ , i.e. the solutions which lie in  $F^n$ . Consequently, we can recover rational points of varieties as fixed points of a set under the action of the Galois group. Sometimes  $Z_f(\bar{F})$  has the structure of an abelian group. In this case we can use techniques from homological algebra to study the action of  $G_F$ . This is done via Galois cohomology. If  $G_F$  acts continuously on an abelian group  $A$ , we can define cohomology groups  $H^q(G_F, A)$ . The abelian group  $H^0(G_F, A)$  is the abelian group of fixed points  $A^{G_F}$ , while  $H^q(G_F, -)$  is the  $q$ th derived functor of the fixed point functor. These cohomology groups are involved in the local Tate duality.

### Local Tate Duality

Let  $F$  be a non-archimedean local field, e.g. a finite extension of  $\mathbb{Q}_p$ . In [6], Tate proves a duality theorem on the Galois cohomology of  $F$ . If  $A$  is an abelian group with a continuous action of  $G_F$ , we define the dual module as

$$A^* := \text{Hom}(A, \bar{F}^\times).$$

This has the structure of a continuous  $G_F$ -module via  $(g \cdot f)(a) := g \cdot f(g^{-1} \cdot a)$ ,  $\forall g \in G, \forall f \in A^*, \forall a \in A$ .

In [6], Tate shows that we have a (cup-product) pairing

$$H^q(G_F, A) \times H^{2-q}(G_F, A^*) \rightarrow \mathbb{Q}/\mathbb{Z} \quad (35.1)$$

for all  $q$ . Moreover, he states that if  $A$  is finite, the cohomology groups are finite and this pairing is perfect. This result is called local Tate duality and it identifies  $H^{2-q}(G_F, A^*)$  with the Pontryagin dual of  $H^q(G_F, A)$  for all  $q$ . This result is very important, but it has some shortcomings.

- Any attempt to enlarge this duality to finitely generated coefficients fails. Indeed, if  $A = \mathbb{Z}$  and  $q = 0$ , (1.1) is the pairing  $\mathbb{Z} \times \mathbb{Q}/\mathbb{Z} \rightarrow \mathbb{Q}/\mathbb{Z}$ , which can't be perfect.
- The topology is not taken into account in this treatment. For example, if  $A = \overline{F}^\times$ , we are considering both  $A$  and  $H^0(G_F, A) = F^\times$  as mere discrete abelian groups. In this way, we lose precious information about their topology.

## Condensed local Tate duality for the Weil group

During my PhD I overcome the two shortcomings presented at the end of the previous section in the case where  $F$  is a  $p$ -adic field, i.e. a finite extension of  $\mathbb{Q}_p$ . I build a "topological" cohomology theory for the Weil group of  $F$ . The cohomology groups are proved to coincide with Galois cohomology groups when the coefficients are finite, and they are proved to be locally compact in some special cases. This allows to enlarge the local Tate duality to a more general class of non-necessarily finite coefficients, where it takes the form of a Pontryagin duality between locally compact abelian groups.

The following results are proved in [1].

### The tools

First of all, following Lichtenbaum's intuitions [3] and inspired by Karpuk's work [2], I replace the Galois group  $G_F$  with its subgroup  $W_F$ , the Weil group.

Secondly, in order to topologise the coefficients and the cohomology groups, I use the theory of Condensed Mathematics [5]. This theory provides a framework where it is possible to do algebra with topological spaces. In particular, locally compact abelian groups embed in condensed abelian group, an abelian category where it is possible to do homological algebra.

### Enlarging local Tate duality as a Pontryagin duality

If  $A$  is a condensed abelian group with an action of the pro-condensed group  $\hat{W}_F$ , I define for all  $q$  condensed abelian groups  $\mathbb{H}^q(\hat{W}_F, A)$ , called condensed Weil cohomology groups. Moreover, I define a dual  $\hat{W}_F$ -module  $A^D$  and I show that there is a cup-product pairing

$$\mathbb{H}^{2-q}(\hat{W}_F, A) \times \mathbb{H}^q(\hat{W}_F, A^D) \rightarrow \mathbb{R}/\mathbb{Z}. \quad (35.2)$$

When  $A$  is finite, this pairing coincides with (1.1) and hence it is perfect by local Tate duality.

I prove that this pairing is perfect for a class of non-necessarily finite coefficients  $A$  (including finite-dimensional real vector spaces, finitely generated discrete abelian groups, finite powers of  $\mathbb{R}/\mathbb{Z}$ ) with an action of a finite quotient of  $G_F$ . In these cases, the cohomology groups  $\mathbb{H}^{2-q}(\hat{W}_F, A)$  and  $\mathbb{H}^q(\hat{W}_F, A^D)$  are locally compact abelian groups. Consequently, (1.2) enlarges Tate duality, expressing it as a Pontryagin duality between locally compact abelian groups.

## Perspectives

Another important duality theorem in arithmetics is the Tate duality for abelian varieties ([4, Corollary 3.4]). If  $\mathcal{A}$  is an abelian variety over a  $p$ -adic field  $F$  and  $\mathcal{A}^t$  is its dual abelian variety, then  $G_F$  acts continuously on  $\mathcal{A}(\overline{F})$  and  $\mathcal{A}^t(\overline{F})$ . Moreover, we have a perfect cup-product pairing for all  $q$

$$H^q(G_F, \mathcal{A}(\overline{F})) \times H^{1-q}(G_F, \mathcal{A}^t(\overline{F})) \rightarrow \mathbb{Q}/\mathbb{Z}.$$

My goal is to rephrase this duality theorem as a Pontryagin duality for the condensed cohomology of the Weil group. Combining this with the results of [1], I could obtain a common generalisation of condensed local Tate duality for finitely generated modules and for abelian varieties, thanks to algebraic objects called 1-motives.

## Bibliographie

- [1] M. Artusa. (2024). "Duality for condensed cohomology of the Weil group of a  $p$ -adic field".
- [2] D. Karpuk. (2013). "Cohomology of the Weil group of a  $p$ -adic field". *Journal of Number Theory* 133.4, 1270–1288.
- [3] S. Lichtenbaum. (2009). "The Weil-étale topology for number rings". *Annals of Mathematics* (2) , 170.2, 657–683.
- [4] J.S. Milne, J. Coates, and S. Helgason. Arithmetic Duality Theorems. Academic Press. Kea books, 1986.
- [5] P. Scholze. Lectures on Condensed Mathematics. 2019.
- [6] J. Tate. (1962). "Duality theorems in Galois cohomology over number fields". *Proc. Int. Congress Math.* Stockholm, 288–295.

# CRYPTOSYSTEMS BASED ON CLASS GROUPS OF IMAGINARY QUADRATIC FIELDS

Agathe BEAUGRAND

Direction : Guilhem CASTAGNOS  
Co-Direction : Fabien LAGUILLAUMIE

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

**E**ncryption based on class groups - especially the CL encryption scheme - presents properties that make it a useful primitive for multiparty computation. However, MPC protocols use zero-knowledge arguments, that are tricky to design for CL statements. We provide some efficient protocols for zero-knowledge in the CL framework, using partial extractability and the  $\mathcal{C}$ -rough assumption.

## Context

### Encryption based on class groups

**Asymmetric encryption.** In asymmetric encryption, every user has a couple of keys  $(pk, sk)$ , with a public key  $pk$  that they can broadcast to anyone wanting to communicate with them, and a secret key  $sk$  that is kept private. The public key allows to encrypt messages, while the secret key is necessary to decrypt ciphertexts. In this context, we consider semantic security : an adversary who knows a ciphertext, and all the public parameters, including the public key, should not be able to learn any information on the encrypted message. To ensure security, encryption schemes rely on hardness assumptions. Such an assumption states that a given computational problem is a hard problem, i.e. there exists no algorithm solving it in polynomial time. An encryption scheme is said secure under a hardness assumption if obtaining information on the message without the secret key is at least as difficult as solving the underlying problem. With well-chosen parameters, an adversary with a polynomial computing capacity will not be able to obtain information. Hardness assumptions come in various flavors, depending on which mathematical objects are considered. Amongst the most famous ones, we can cite the factorization problem over integers, the discrete logarithm problem over groups, or the short vector problem over lattices.

**Class groups in cryptography.** Since the late 80s, class groups of quadratic fields have been used as building blocks in cryptography, following the development of techniques that allow to perform computation efficiently over these groups. In addition to having a hard to compute discrete logarithm, these groups are said "of unknown order", as the best existing algorithms for computing a class group's order have sub exponential complexity. In particular, computing a class number is considered harder than the factorization problem or discrete logarithms in multiplicative groups of finite

fields, which means that the parameters can be chosen smaller for class groups based encryption than they are for RSA or Paillier.

**The CL cryptosystem.** The CL encryption scheme is a discrete-logarithm based asymmetric cryptosystem, proposed by Castagnos and Laguillaumie in [3]. It is built upon groups of unknown orders, and more particularly class groups of imaginary quadratic fields. Concretely, we construct a cyclic group  $G$  of unknown order, and sample a generator  $g$  of the subgroup  $G^q$  of  $q$ -th powers. We also construct a subgroup  $\langle f \rangle = F \subset G$  of prime order  $q$ , in which the discrete logarithm is easily computable. The group  $G$  is such that  $G \cong G^q \times F$ . The secret key is an integer  $x \in \mathbb{Z}$ , and the public key is  $h = g^x$ . The CL ciphertext for a message  $m \in \mathbb{Z}_q$  with randomness  $r \in \mathbb{Z}_q$  is

$$c = (g^r, h^r f^m).$$

This encryption scheme is linearly homomorphic, meaning that from encryptions of messages  $m$  and  $m'$ , one can compute an encryption of any  $am + m'$ , without knowing the secret key.

### Multiparty computation (MPC)

Multiparty computation allows two or more parties to perform computations together while preserving their privacy. More precisely, each party  $P_i$  has a secret input  $x_i$ , and  $P_1, \dots, P_n$  want to compute a function  $f(x_1, \dots, x_n)$  without revealing anything about their input more than what can be learned from the value of  $f$ . A useful tool for performing multiparty computation is homomorphic encryption, which allows to perform some of the computation directly inside ciphertexts, without having to decrypt. The CL scheme, thanks to its linear homomorphism, and exploiting the fact that the space of messages is  $\mathbb{Z}_q$ , has proved a very interesting primitive for multiparty computation.

Aside from privacy, MPC needs to guarantee the result of the computation, and avoid misbehaviours from the parties. Once again, CL presents an advantage for

this matter, as its parameters can be computed in a transparent manner. This way, the correct generation of parameters is guaranteed without having to rely on an external trusted party.

To ensure correctness, MPC protocols also make extensive use of zero-knowledge proofs and arguments. Zero-knowledge arguments of knowledge (ZKAoK) are two-party protocols between a prover and a verifier designed to allow the prover to convince the verifier that he knows a specific value, and can be used for him to prove that he performed a computation honestly.

To exploit CL's advantages in MPC, we need efficient zero-knowledge arguments in the CL framework. However, designing such arguments raises technical challenges due to the unknown order of the underlying group.

## ZKAoKs for CL statements

### ZKAoKs for CL : a technical challenge

**Schnorr in a group of prime order.** In Schnorr protocol, the prover tries to convince the verifier that he knows the discrete logarithm  $x$  of a group element  $h$  in base  $g$ . It consists in 4 steps :

1. The prover sends  $\tilde{h} = g^a$ , for  $a \in \mathbb{Z}_q$  random.
2. The verifier sends a random challenge  $b \in \mathbb{Z}_q$ .
3. The prover answers with  $y = a + bx$ .
4. The verifier accepts the proof if  $\tilde{h}.h^b = g^y$ .

If the prover is able to answer correctly for a random challenge, we can assume he can answer for two different challenges  $b$  and  $b'$ . In this case,  $g^{y-y'} = h^{b-b'}$ ,  $\delta := b - b' \neq 0$  is invertible modulo the order of the group, so that  $h = g^{(y-y')\delta^{-1}}$ , and the discrete logarithm can be extracted from accepted proofs, by computing  $(y - y')\delta^{-1}$ . The protocol is said **knowledge sound**.

**Switching to CL statements.** In the CL framework, achieving knowledge soundness is more technical : as the order of the group is unknown, we cannot check that  $\delta$  is invertible, let alone compute an inverse. However, if  $\delta$  is invertible, the equality  $g^{y-y'} = h^\delta$  still guarantees that  $h \in \langle g \rangle$ . Some previous solutions for obtaining knowledge soundness used challenges in  $\{0, 1\}$  [2], or relied on the  $\mathcal{C}$ -rough assumption [1]. This assumption states that, given a class group, one cannot decide whether or not it has subgroups of order smaller than  $\mathcal{C}$ . This assumption allows to assume that, by default, the considered class number does not have small factors.

### Contributions

**Partial extractability.** To improve the efficiency of proofs for CL statements, we exploit the structure  $G = G_q \times F$ . Indeed, a closer look at CL ciphertexts shows that although the group has unknown order, the meaningful information, i.e. the message, is encoded in a subgroup of prime order, while the unknown order part only contains the randomness. But, most arguments on ciphertexts - such as arguments of correct

generation, application of homomorphic operations, correct decryption - aim at proving that the ciphertexts can be decrypted to the expected message, but do not need to specify the randomness. We thus define **partial extractability**, as an intermediary soundness notion, allowing to only extract a part of the witness.

**Efficient and succinct arguments.** Using the  $\mathcal{C}$ -rough order assumption, we design a batched argument for correctness of CL ciphertexts with partial extractability on the messages, and a constant communication cost. A batched argument allows to prove many similar statements in a unique argument, in a way that's more efficient than running a single argument for each statement.

We also focus on the design of efficient shuffle arguments. A shuffle is an operation consisting in mixing and rerandomizing a set of ciphertexts, and has applications in MPC protocols such as mixnets or private set intersection. We propose two protocols for an argument of shuffle. The first one has a communication sublinear in the number of ciphertexts and a constant number of rounds, while the second, based on bulletproofs techniques [4], has logarithmic communication, at the expense of a logarithmic number of rounds.

## Bibliographie

- [1] L. Braun, I. Damgård, and C. Orlandi. Secure multiparty computation from threshold encryption based on class groups. In *Advances in Cryptology - CRYPTO 2023*, pages 613-645, Cham, 2023. Springer Nature Switzerland.
- [2] G. Castagnos, D. Catalano, F. Laguillaumie, F. Savasta, and I. Tucker. Two-party ECDSA from hash proof systems and efficient instantiations. In *Advances in Cryptology - CRYPTO 2019, Part III*, volume 11694 of *Lecture Notes in Computer Science*, pages 191-221. Springer
- [3] G. Castagnos and F. Laguillaumie. Linearly homomorphic encryption from DDH. In *Topics in Cryptology - CTRSA 2015*, volume 9048 of *Lecture Notes in Computer Science*, pages 487- 505. Springer
- [4] M. Hoffmann, M. Klooß, and A. Rupp. Efficient zero-knowledge arguments in the discrete log setting, revisited. In *ACM CCS 2019 : 26th Conference on Computer and Communications Security*, pages 2093-2110. ACM Press



# NON-CONVEX METHODS FOR INVERSE PROBLEMS WITH LOW DIMENSIONAL MODELS

Pierre-Jean BÉNARD

Direction : Yann TRAONMILIN

Co-Direction : Jean-François AUJOL

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

Dans cette thèse nous proposons d'étudier la reconstruction de signaux parcimonieux sur le domaine continu à l'aide d'une acquisition compressée composée de mesures linéaires dans le cadre des problèmes inverses. Bien que de nombreux algorithmes et leurs études théoriques associées existent dans ce domaine, leur temps d'exécution en pratique est d'autant plus ralenti que la taille des signaux à reconstruire est grande. Nous proposons une nouvelle méthode basée sur l'état de l'art qui élimine ce ralentissement. Nous fournissons également des garanties théoriques en lien avec ce nouvel algorithme, quelques applications pratiques ainsi que des axes d'accélération supplémentaires.

## Contexte

La reconstruction de signaux parcimonieux en continu est un problème classique dans le champs des problèmes inverses avec de nombreuses applications. Nous posons  $x_0$  comme le signal inconnu à reconstruire qui l'on modélise comme une somme de  $K$  mesures de Dirac sur  $\mathbb{R}^d$  :

$$x_0 = \sum_{i=1}^K a_i \delta_{t_i} \quad (37.1)$$

avec  $a \in \mathbb{R}^K$  les amplitudes et  $t \in \mathbb{R}^{K \times d}$  les positions des impulsions. En ce basant sur ce modèle, les seules informations nécessaires pour reconstruire  $x_0$  sont ses amplitudes  $a$  et positions  $t$  associées. L'observation  $y \in \mathbb{C}^m$  correspond à l'acquisition compressée composée de  $m$  mesures linéaires. Elle peut être modélisée dans le cas sans bruit par

$$y = Ax_0 \quad (37.2)$$

avec  $A$  l'opérateur linéaire qui fait correspondre l'espace  $\mathcal{M}$  des mesures finies et signées sur  $\mathbb{R}^d$  dans  $\mathbb{C}^m$ . L'opérateur linéaire  $A$  modélise l'acquisition, il peut représenter par exemple un échantillonnage de  $m$  mesures de Fourier ou l'intensité de chaque pixel d'une image prise au microscope. L'une des manière de reconstruire  $x_0$  à partir de  $y$  et  $A$  est de trouver le minimiseur d'un problème des moindres carrés non-convexe :

$$\mathcal{P} : x^* \in \operatorname{argmin}_{x \in \Sigma_{K,\epsilon}} \|y - Ax\|_2^2 \quad (37.3)$$

avec  $\Sigma_{K,\epsilon}$  l'ensemble des signaux de  $K$  impulsions avec une contrainte de séparation entre les impulsions. Des garanties théoriques de reconstruction ont été données dans [1] pour des opérateurs d'acquisition  $A$  ayant une propriété d'isométrie restreinte (RIP). D'autres résultats théoriques s'articulent autour de ce problème et expliquent en partie le succès d'une méthode gloutonne : le Sliding Continuous Orthogonal Matching Pursuit (Sliding COMP). À noter que  $\mathcal{P}$  n'est pas la seule formulation du problème capable de reconstruire  $x_0$ , on peut

citer sa version relaxée (et donc convexe) qui impose un contrainte sur le nombre d'impulsions sur le signal estimé. Un algorithme associé est le Sliding Frank-Wolfe, méthode similaire à Sliding COMP dans leur structure. Ces deux méthodes agissent comme une boucle qui à chaque itération ajoute une impulsion au signal estimé basé sur sa corrélation avec le résidu entre l'observation  $y$  et l'estimation actuelle. Puis à la fin de chaque itération, elles effectuent une descente sur l'ensemble des paramètres  $(a, t)$  du signal estimé. Cette descente (l'étape la plus coûteuse des deux méthodes) est de plus en plus ralentie au fur et à mesure que le nombre d'impulsions estimées augmente. C'est pourquoi nous avons introduit la méthode COMP sur-paramétrisé avec descente de gradient projetée (OP-COMP + PGD), basée sur Sliding COMP, pour palier à ce problème.

## COMP sur-paramétrisé avec une descente de gradient projetée

### Introduction à la méthode et résultats théoriques

Notre méthode OP-COMP + PGD (introduite dans [2]) est une modification de la méthode état de l'art Sliding COMP. Elle peut être décomposée en deux parties distinctes. La première partie (OP-COMP) est équivalente à Sliding COMP, à la seule différence que nous supprimons la descente sur tous les paramètres à chaque itération. En sur-paramétrisant le signal, c'est à dire en ajoutant plus d'impulsions que nécessaire, nous compensons le manque de descente. En deuxième partie (PGD), elle effectue une unique descente sur tous les paramètres une fois toutes les itérations de OP-COMP terminées. Afin de réduire le nombre d'impulsions du signal estimé, elle procède à une projection heuristique qui fusionne les impulsions trop proches l'une de l'autre. Cela impose au signal estimé de rester dans notre ensemble des solutions  $\Sigma_{K,\epsilon}$ .

Nous accompagnons ce nouvel algorithme avec une

étude théorique sur OP-COMP [3] indiquant alors que, pour un opérateur linéaire  $A$  suivant une RIP, les  $K$  premières itérations de OP-COMP, les impulsions estimées peuvent être initialisées arbitrairement proches des impulsions du signal à reconstruire  $x_0$ . En combinant ce résultat avec des travaux portant sur la taille des bassins d'attraction du problème des moindres carrés non-convexe  $\mathcal{P}$ , nous pouvons alors assurer qu'avec une bonne initialisation du signal estimé, ce dernier se trouve dans le bassin d'attraction du minimum global et donc OP-COMP + PGD peut reconstruire  $x_0$  exactement.

### Résultats expérimentaux

En termes de résultats expérimentaux, nous avons comparé OP-COMP + PGD à Sliding COMP sur plusieurs configurations : différents opérateurs linéaires tels que l'échantillonnage de Fourier et MA-TIRF en microscopie et avec un nombre d'impulsions par signaux à reconstruire variant de 5 à 100. En récapitulatif, la qualité de reconstruction par OP-COMP + PGD est identique à celle de Sliding COMP. En terme de temps de calcul, pour des très petits nombre d'impulsions  $K$  (inférieur à 15), Sliding COMP reste plus rapide mais au-delà, OP-COMP devient moins coûteux avec un temps total d'ordre  $\mathcal{O}(K)$  contre  $\mathcal{O}(K^2)$  pour Sliding COMP.

### Méthodes d'accélération

En ajout de OP-COMP + PGD, nous avons également mis au point d'autres techniques d'accélération qui peuvent se combiner l'une à l'autre.

#### La descente en coordonnées par blocs

L'idée derrière la descente en coordonnées par blocs profite de la parcimonie de notre modèle. Cette méthode cherche à réduire les coûts de calculs lors de la descente de gradient projetée [4]. Elle agit ainsi en découpant le signal estimé en différent blocs en fonction d'un critère similaire à celui de Gauss-Southwell. Ce critère examine chaque impulsion  $i \in \{1, \dots, K\}$  séparément et détermine si cette dernière a convergé ou non en regardant la norme du gradient de la fonctionnelle associé aux paramètres  $(a_i, t_i)$ . Ensuite, l'ensemble des impulsions qui ont déjà convergé sont mises de côté et l'on effectue une descente de gradient uniquement sur les impulsions qui n'ont pas encore convergé. En répétant ce processus jusqu'à convergence du signal estimé, nous réduisons les calculs impliqués dans la descente des paramètres. En comparant cette méthode à OP-COMP + PGD sans descente en coordonnées par blocs, nous observons une amélioration en terme de temps de calcul total de 35% sans perte de précision dans la reconstruction de  $x_0$ .

#### L'esquisse de l'opérateur linéaire

Dans cette section, nous étudions l'échantillonnage d'opérateur linéaire. Dans certains cas, les calculs impliquant l'opérateur linéaire peuvent être très coûteux dû, soit au grand nombre de mesures, soit à l'expression

et implémentation de l'opérateur linéaire. Pour chercher à réduire le nombre de mesures et/ou simplifier l'expression de  $A$ , nous effectuons un échantillonnage dans l'espace des fréquences avec une transformée de Fourier [5]. Nous obtenons alors la transformée exacte d'un opérateur  $\mathcal{S}A^{id}$  et la transformée discrète  $\mathcal{S}_dy$  de notre observation  $y$ . En résolvant un problème similaire à  $\mathcal{P}$  avec nos nouvelles données, nous pouvons alors reconstruire un signal approximant  $x^*$  et dont la précision avec  $x_0$  dépend du nombre de fréquence  $m_s$  choisi. En comparant cette méthode avec OP-COMP + PGD sans échantillonnage, nous obtenons une précision comparable à la méthode sans échantillonnage dès que l'on prend un nombre de fréquence  $m_s$  supérieur à 15% du nombre de mesures total de  $A$ . Concernant le temps de calcul total, à 15% d'échantillonnage, nous obtenons une accélération de 80% par rapport à la méthode sans échantillonnage.

### Remerciements

Ce travail est soutenu par l'Agence Nationale de la Recherche (ANR) sous la référence ANR-20-CE40-0001 (projet EFFIREG).

### Bibliographie

- [1] Gribonval, R., Blanchard, G., Keriven, N., and Traonmilin, Y. (2021). Compressive Statistical Learning with Random Feature Moments. *Mathematical Statistics and Learning*, 3(2) :113–164.
- [2] Bénard, P.-J., Traonmilin, Y., and Aujol, J.-F. (2022). Fast off-the-grid sparse recovery with over-parametrized projected gradient descent. In *30th EUSIPCO*, 2206–2210.
- [3] Bénard, P.-J., Traonmilin, Y., Aujol, J.-F., and Soubiès E. (2024). Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent : theory and application. In *Inverse Problems*, 40(5) :055010.
- [4] Bénard, P.-J., Traonmilin, Y., and Aujol, J.-F. (2024). Projected Block Coordinate Descent for sparse spike estimation. In *preprint*.
- [5] Bénard, P.-J., Traonmilin, Y., and Aujol, J.-F. (2024). Sketched Over-Parametrized Continuous Orthogonal Matching Pursuit for sparse spike estimation. In *preprint*.

# COUPLAGES DE MOUVEMENTS BROWNIENS EN GÉOMÉTRIE SOUS-RIEMANNIENNE

Magalie BÉNÉFICE

Direction : Michel BONNEFONT  
Co-Direction : Marc ARNAUDON

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

L'objectif de ma thèse est l'étude des couplages de mouvements browniens en géométrie sous-riemannienne, plus particulièrement sur  $SU(2)$ ,  $SL(2, \mathbb{R})$  et sur les groupes de Carnot libres d'ordre 2. Les couplages ont de nombreuses applications en géométrie, en transport optimal ainsi qu'en analyse. Leur utilisation est en particulier très pertinente en géométrie sous-riemannienne car on peut alors se passer de certaines notions difficiles à définir et à utiliser (la courbure par exemple). Le but de cette thèse est d'utiliser cette approche pour obtenir de nouveaux résultats ou de nouvelles preuves tels que des inégalités de gradients pour le semi-groupe de la chaleur ou des propriétés sur les fonctions harmoniques.

## Contexte d'étude

### Notion de couplage

Nous donnons tout d'abord une définition de la notion de couplage. Soit  $G$  une variété et  $g, \tilde{g} \in G$ . On appelle couplage de deux mouvements browniens partant de  $(g, \tilde{g})$  toute mesure de probabilité sur  $\mathcal{C}([0, T], G \times G)$  (où  $T \in [0, +\infty]$ ) pour laquelle la première marginale suit la loi d'un mouvement Brownien  $(\mathbb{B}_t)_t$  partant de  $g$  et la seconde marginale suit la loi d'un mouvement brownien  $(\tilde{\mathbb{B}}_t)_t$  partant de  $\tilde{g}$ . En d'autres termes, il s'agit d'étudier la loi jointe du processus  $(\mathbb{B}_t, \tilde{\mathbb{B}}_t)_t$ .

Donnons quelques exemples simples (mais non moins utiles) de couplages partant de  $(g, \tilde{g})$  sur  $\mathbb{R}^2$ .

1. **Le couplage synchrone** : On construit  $\mathbb{B}_t$  et  $\tilde{\mathbb{B}}_t$  égaux à une translation près :  $\tilde{\mathbb{B}}_t - \tilde{g} = \mathbb{B}_t - g$ .
2. **Le couplage par réflexion** : On choisit une base orthonormée  $(e_1, e_2)$  de  $\mathbb{R}^2$  telle que  $e_1 := \frac{g - \tilde{g}}{\|g - \tilde{g}\|_2}$ . Soit  $(B_t^1, B_t^2)_t$  un mouvement brownien 2-dimensionnel partant de 0. On construit alors :

$$\mathbb{B}_t = g + B_t^1 e_1 + B_t^2 e_2 \text{ et } \tilde{\mathbb{B}}_t = \tilde{g} - B_t^1 e_1 + B_t^2 e_2.$$

Les deux mouvements browniens  $(\mathbb{B}_t)_t$  et  $(\tilde{\mathbb{B}}_t)_t$  sont en fait construits grâce à une symétrie axiale.

Notons que ces deux couplages sont **co-adaptés**, c'est à dire que le futur des deux processus construits ne dépend que de leur passé commun. En fait  $(\mathbb{B}_t, \tilde{\mathbb{B}}_t)_t$  est même markovien. Nous proposons un troisième exemple qui n'est pas co-adapté :

3. **Un couplage par décomposition** : Soit  $T > 0$ , on peut décomposer le mouvement brownien dans une base de  $L^2([0, T])$ . Par exemple, en utilisant

la décomposition de Karhunen Loève, on a :

$$\begin{aligned} \mathbb{B}_t &= g + \sqrt{T} \sum_{j \geq 0} \xi_j \frac{\sqrt{2}}{j\pi} \sin\left(\frac{j\pi t}{T}\right) \\ \tilde{\mathbb{B}}_t &= \tilde{g} + \sqrt{T} \sum_{j \geq 0} \tilde{\xi}_j \frac{\sqrt{2}}{j\pi} \sin\left(\frac{j\pi t}{T}\right) \end{aligned}$$

où  $(\xi_j)_j$  (resp.  $(\tilde{\xi}_j)_j$ ) sont identiquement et indépendamment distribuées et suivent une loi normale  $\mathcal{N}(0, 1)$ . On peut alors construire le couplage  $(\mathbb{B}_t, \tilde{\mathbb{B}}_t)_t$  en choisissant un couplage  $(\xi_j, \tilde{\xi}_j)$  pour tout  $j \geq 0$ . En prenant par exemple  $\xi_1 \neq \tilde{\xi}_1$  et  $\xi_0 = \tilde{\xi}_0$ , on a  $\tilde{\mathbb{B}}_T = \mathbb{B}_T - g + \tilde{g}$  : le couplage n'est pas co-adapté puisqu'on a besoin de la connaissance du futur (la valeur  $\mathbb{B}_T$ ) de l'un des processus pour construire le second.

Ces exemples se généralisent à  $\mathbb{R}^n$  en toute dimension ainsi qu'à certaines variétés riemanniennes comme la sphère ou le plan hyperbolique.

### Temps de couplage

Pour un couplage  $(X_t, \tilde{X}_t)_t$  donné, il peut être intéressant d'étudier le premier temps de couplage  $\tau$ , c'est à dire le premier temps de rencontre des mouvements browniens. En effet, il est aisé de montrer l'inégalité de couplage suivante :

$$\mathbb{P}(\tau > s) \geq d_{TV}(\mathcal{L}(X_s), \mathcal{L}(\tilde{X}_s)) \quad \forall s > 0$$

où  $d_{TV}$  est la distance en variation totale. En particulier, les couplages pour lesquels cette inégalité devient une égalité sont appelés **couplage maximaux**. Si  $\mathbb{P}(\tau > t)$  et  $d_{TV}(\mathcal{L}(X_t), \mathcal{L}(\tilde{X}_t))$  ont le même ordre de grandeur pour  $t$  grand, le couplage est dit **efficace**. Dans le cas de variétés riemanniennes qui ont de bonnes propriétés de symétrie, ces couplages sont obtenus avec un couplage par réflexion. En étudiant la quantité  $\mathbb{P}(\tau > t)$ , on peut

aussi obtenir des inégalités pour le gradient du semi groupe de la chaleur et des résultats sur les fonctions harmoniques.

### Mouvements Browniens en géométrie sous-riemannienne

Sur des variétés sous-riemanniennes, les mouvements browniens peuvent souvent s'écrire sous la forme  $\mathbb{B}_t = (X_t, z_t)$  avec  $(X_t)_t$  un processus de diffusion elliptique et  $z_t = f(X_s)_{s \leq t}$  où  $f$  est une fonctionnelle. Grâce à cette structure, il suffit de construire un couplage de processus elliptiques (bien mieux connus)  $(X_t, \tilde{X}_t)_t$  pour obtenir un couplage de mouvements Browniens. Durant ma thèse j'ai étudié les couplages de mouvements browniens sur les variétés sous-riemanniennes suivantes :

- Le groupe d'Heisenberg  $\mathbb{H}$ , pour lequel  $X_t$  est un mouvement brownien sur  $\mathbb{R}^2$  et  $z_t$  est l'aire signée (à valeurs dans  $\mathbb{R}$ ) balayée par  $(X_s)_{s \leq t}$ .
- Le groupe spécial unitaire  $SU(2)$ , où  $X_t$  est un mouvement brownien sur la sphère  $S^2$  et où  $z_t$  est l'aire signée (modulo  $4\pi$ ) balayée par  $(X_s)_{s \leq t}$ .
- Le groupe spécial linéaire  $SL(2, \mathbb{R})$ , où  $X_t$  est un mouvement brownien sur le plan hyperbolique et où  $z_t$  est l'aire signée (modulo  $4\pi$ ) balayée par  $(X_s)_{s \leq t}$ .

Ces trois espaces peuvent être vus comme les espaces modèles de courbure constante en géométrie sous-riemannienne.

Plus récemment, je me suis aussi intéressée aux groupes de Carnot libres d'ordre 2 et de dimension  $n \geq 3$ . Dans ce cas,  $X_t$  est un mouvement brownien sur  $\mathbb{R}^n$  et  $z_t = (z_t^{i,j})_{1 \leq i < j \leq n} \in \mathbb{R}^{\frac{n(n-1)}{2}}$  où  $z_t^{i,j}$  est l'aire signée (à valeurs dans  $\mathbb{R}$ ) balayée par le processus  $(X_s^i, X_s^j)_{s \leq t}$ .

### Résultats

J'ai majoritairement étudié les couplages avec succès sur les espaces susmentionnés. Dans le cas du groupe d'Heisenberg nous pouvons citer deux constructions notables (à quelques variantes près). Le premier est un couplage avec succès co-adapté proposé par Kendall [1]. Cette construction utilise en alternance les couplages miroirs et synchrone. Notons que ce couplage a été généralisé par Kendall [1] aux groupes de Carnot libres d'ordre 2. Le second est un couplage non co-adapté proposé par Banerjee, Gordina et Mariano [2]. Ce couplage est construit en se ramenant à deux mouvements browniens partant de la même fibre et en utilisant le couplage par décomposition. On choisit alors  $\xi_k = \tilde{\xi}_k$  pour tout  $k \neq 1$ . En particulier il a été prouvé que ce couplage est efficace et que ce critère d'efficacité ne peut être atteint avec un couplage co-adapté (ce critère inclut le couplage de Kendall). On obtient :

$$\mathbb{P}(\tau > t) \leq C \left( \frac{\|X_0 - X'_0\|_2}{\sqrt{t}} + \frac{|z_0 - z'_0 + X_0^1 X_0^{2'} - X_0^{1'} X_0^2|}{t} \right).$$

Ce résultat permet d'obtenir des inégalités de gradients. En particulier, pour toute fonction positive harmonique sur un domaine  $D$ , il existe une constante  $C$  telle que pour  $x \in D$  et  $\delta_x = d(x, D^c)$ , on obtient :

$$\|\nabla u(x)\| < C \left( 1 + \frac{1}{\delta_x} + \frac{1}{\delta_x^4} + \frac{(1 + \delta_x)^3}{\delta_x^4} \right) u(x).$$

J'ai pu généraliser ces deux couplages à  $SU(2)$  dans [3, 4]. En considérant des mouvements browniens partant de la même fibre ( $X_0 = \tilde{X}_0$ ), le couplage non co-adapté se généralise aussi à  $SL(2, \mathbb{R})$ . Une des difficultés majeures de ce travail est la prise en compte de la courbure de la sphère dans laquelle vit  $(X_t, \tilde{X}_t)_t$ . Sur  $SU(2)$ , en notant  $d_{cc}$  la distance sous-riemannienne usuelle et pour  $t$  suffisamment grand, le couplage non co-adapté entraîne l'inégalité suivante :

$$\mathbb{P}(\tau > t) \leq C e^{-ct} d_{cc}((X_0, z_0), (X'_0, z'_0))$$

où  $C$  et  $c$  sont des constantes indépendantes de la distance initiale (pour  $SL(2, \mathbb{R})$  cette inégalité reste vraie pourvu que  $X_0 = X'_0$ ). On peut déduire de ce résultat une inégalité de gradient sur  $SU(2)$ . Pour  $t$  suffisamment grand :

$$\|\nabla_{\mathcal{H}} P_t f(g)\|_{\mathcal{H}} \leq 2\|f\|_{\infty} C e^{-ct} \text{ p.s..}$$

Sur  $SL(2, \mathbb{R})$ , on obtient un argument direct pour prouver que toute fonction harmonique bornée est constante sur chaque fibre.

La généralisation du couplage non co-adapté et de toutes ses applications au cas des groupes de Carnot libre d'ordre 2 en toute dimension est un travail en cours. La difficulté majeure vient du fait de devoir travailler sur une des aires balayées  $(z_t^{i,j})_t$  tout en gardant un certain contrôle sur les autres.

### Bibliographie

- [1] Kendall, W. (2007). Coupling all the Lévy stochastic areas of multidimensional Brownian motion. Ann. Probab., 35, no.3, 935-953.
- [2] Banerjee, S., Gordina, M., Mariano, P. (2018). Coupling in the Heisenberg group and its applications to gradient estimates. Ann. Probab. 46, no.6, 3275-3312.
- [3] Bénéfice, M. (2023) Non co-adapted couplings of Brownian motions on subRiemannian manifolds. <https://arxiv.org/abs/2312.14512>.
- [4] Bénéfice, M. (2023). Non co-adapted couplings of Brownian motions on subRiemannian manifolds. <https://arxiv.org/abs/2305.10017>.



# ANALYSE ET DÉVELOPPEMENT D'ALGORITHMES D'OPTIMISATION POUR LES RÉSEAUX DE NEURONES

**Bilel BENSAID**

**Direction : Rodolphe TURPAULT**

**Co-Direction : Gaël POETTE**

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

L'objectif de cette thèse est d'analyser les algorithmes d'optimisation intervenant dans l'entraînement des réseaux de neurones. Ces derniers sont extrêmement sensibles aux réglages d'un certain nombre d'hyperparamètres et souffrent d'un manque de garantie théorique sous des hypothèses générales de non-convexité. Notre approche permet à la fois de rendre ces algorithmes adaptatifs et de donner des garanties de convergence pour les optimiseurs les plus populaires.

## Stabilité de Lyapunov

### Instabilités et ODEs

Pour des choix d'hyperparamètres assurant empiriquement la convergence vers un point critique de l'erreur  $\mathcal{R}(\theta)$  du réseau ( $\theta$  les poids), les optimiseurs classiques (p-GD, Momentum, RMSProp, Adam) présentent des instabilités : initialisés arbitrairement proche d'un minimum global, ils peuvent converger vers un minimum local de piètre performance [1]. Ce phénomène disparaît quand le taux d'apprentissage (pas de temps :  $\eta$ ) se rapproche de 0. Il est alors naturel de s'intéresser au système asymptotique.

A la limite  $\eta \rightarrow 0$ , ces algorithmes peuvent être approchés par une famille d'EDO. Ces dernières admettent des fonctions de Lyapunov, c'est-à-dire, des fonctionnelles qui décroissent le long des trajectoires solutions. L'existence de telles fonctions permet d'établir la stabilité et l'attractivité de ces EDOs.

### Schéma d'approximation stable

Dès lors, les optimiseurs du Deep Learning sont stables et attractifs asymptotiquement quand le pas de temps est arbitrairement proche de 0. Par conséquent, cela signifie que l'algorithme, c'est-à-dire, la discrétisation des EDOs susmentionnées ne préserve pas la stabilité continue. Considérons une ODE  $y'(t) = F(y(t))$  sur  $\mathbb{R}^m$  stable au voisinage de ses équilibres admettant une fonction de Lyapunov  $V : \mathbb{R}^m \mapsto \mathbb{R}_+$  et notons  $\dot{V}$  sa dissipation :

$$\dot{V}(y) := \nabla V(y) \cdot F(y) \leq 0.$$

Une manière d'espérer la conservation de la stabilité consiste à discrétiser l'EDO de sorte à ce que  $V(y_n)$  décroisse à la même vitesse que  $\dot{V}(y_n)$ , ce qui s'écrit avec un schéma d'Euler explicite ( $\eta_n$  pas de temps adaptatif) :

$$\begin{cases} y_{n+1} = y_n + \eta_n F(y_n), \\ V(y_{n+1}) - V(y_n) \leq -\lambda \eta_n \dot{V}(y_n), \end{cases} \quad (39.1)$$

pour  $\lambda \in ]0, 1[$ . La condition sur  $V$  peut-être assurée algorithmiquement par une recherche linéaire et généralise la fameuse règle d'Armijo. Nous prouvons qu'un tel schéma assure la stabilité quels que soient les hyperparamètres mis en jeu ainsi que l'attractivité [2] :  $d(y_n, \mathcal{Z}) \rightarrow 0$ , où

$$\mathcal{Z} = \left\{ y \in \mathbb{R}^m, \dot{V}(y) = 0 \right\}.$$

Ces résultats peuvent être vus comme une généralisation du théorème de Lyapunov et de LaSalle à des systèmes discrets discontinus étant donné que la recherche linéaire n'est pas un procédé continu. Ce nouveau type d'optimiseur basé sur le contrôle de Lyapunov [3] présente d'excellentes performances sur des réseaux feed-forward et convolutionnels dans le cadre déterministe.

Sur la figure 39.1, les 4 premières lignes retranscrivent les meilleures performances connues pour la reconnaissance des images FASHION-MNIST : les réseaux correspondants possèdent des millions de paramètres et les hyperparamètres sont optimisés par une "gridsearch". Notre nouvel optimiseur (dernière ligne) appliqué sur un réseau convolutionnel à seulement 3000 paramètres dépasse ceux qui étaient obtenus avec des modèles volumineux.

| Archs             | Perfs(%) | Params | Réglage        |
|-------------------|----------|--------|----------------|
| VGG8B             | 95.47    | 7.3M   | Manual         |
| Resnet110         | 95.99    | 1.7M   | Manual         |
| WRN28-10          | 96.35    | 36.5M  | Manual         |
| <b>Fine-DARTS</b> | 96.91    | 3.2M   | Gradient Based |
| <b>LeNet1</b>     | 99.1     | 3.2K   | None           |

TABLE 39.1 – Meilleures réseaux sur la base de donnée FASHION-MNIST



## Convergence globale

La règle imposée sur le pas de temps (1.1) transforme les différents optimiseurs en des systèmes discrets dissipatifs. Pour les ODEs dissipatifs, il est possible d'assurer la convergence des solutions vers un point stationnaire pour les fonctions dites de Kurdyka-Lojasiewicz [4], c'est-à-dire vérifiant l'inégalité suivante au voisinage de chaque point critique  $\theta^*$  :

$$\|\nabla \mathcal{R}(\theta)\| \geq \phi(\mathcal{R}(\theta) - \mathcal{R}(\theta^*)),$$

pour  $\phi$  une fonction concave définie dans le voisinage de 0. Ces fonctions comprennent une grande classe de fonctions non-convexes, dont les fonctions semi-algébriques et semi-analytiques [5] qui englobent la plupart des surfaces de réseaux de neurones.

Dans ce cadre, nous montrons que la plupart des optimiseurs du Deep Learning (p-GD, Momentum, RMSProp) agrémentés de notre règle d'Armijo convergent globalement vers un point critique de  $\mathcal{R}$  tout en explicitant les taux de convergence locales sur cette classe de fonctions. En particulier, nous établissons la première preuve de convergence de RMSProp déterministe à pas constant sans supposer que le gradient est bornée.

## Accélération pour les réseaux profonds

L'approche précédente ne permet pas d'établir une accélération de convergence pour les méthodes inertiels [6]. Cependant, les méthodes de type Adam et gradient normalisé semblent mieux se comporter sur les réseaux très profonds de type RNN/LSTM comparées à de la descente de gradient. Il est vérifié empiriquement que ce type d'architecture satisfait une hypothèse dite de  $(L_0, L_1)$ -régularité :

$$\|\nabla^2 \mathcal{R}(\theta)\| \leq L_0 + L_1 \|\nabla \mathcal{R}(\theta)\|.$$

Il est prouvé que la descente de gradient subit une décélération sur ce type de fonctions [7] due à une trop grande sensibilité à la condition initiale. Nous avons établi que notre algorithme adaptatif ne subit pas cet effet et obtenu une borne moins sensible à la raideur du problème comparée à d'autres techniques d'adaptivité comme le gradient clipping.

## Le cadre mini-batch

Dans le cas où les données ne peuvent entrer en mémoire, la fonction de coût est découpée en plusieurs

groupes appelés batches :  $\mathcal{R} = \sum_{i=1}^m \mathcal{R}_i$ . L'approche couramment utilisée (SGD, Momentum stochastique,...) consiste à tirer uniformément un groupe  $i$  et à appliquer une itération classique de l'optimiseur qu'à partir de l'information sur  $\nabla \mathcal{R}_i$ . Nous avons constaté qu'une telle approche (à pas constant) ne permet pas de minimiser la fonction somme  $\mathcal{R}$  sans l'hypothèse d'interpolation en chaque point critique  $\theta^*$  :

$$\nabla \mathcal{R}(\theta^*) = 0 \implies \forall 1 \leq i \leq m, \nabla \mathcal{R}_i(\theta^*) = 0.$$

Notre approche pour palier à ce problème consiste à interpréter le gradient stochastique comme un splitting d'opérateur (Lie-Trotter) et à construire un splitting dit re-balancé afin de traiter le cas non-interpolant avec des méthodes adaptatives. Cette méthode a des liens profonds avec les récentes méthodes de réduction de variance (SVRG, SARAH,...).

## Bibliographie

- [1] B.Bensaid, G.Poette, R.Turpault (2023). Deep Learning Optimization from a Continuous and Energy Point of View,. Journal of Scientific computing.
- [2] B.Bensaid, G.Poette, R.Turpault (2024). An Abstract Lyapunov Control Optimizer : Local Stabilization and Global Convergence. Submitted.
- [3] I.Karafyllis and L.Grüne (2011). Feedback Stabilization Methods for the Numerical Solution of Ordinary Differential Equations. Discrete and Continuous Dynamical Systems.
- [4] A.Haraux, M.Jendoubi (2015). The Convergence Problem for Dissipative Autonomous Systems : Classical Methods and Recent Advances. Springer.
- [5] J.Bolte, A.Daniilidis, A.Lewis (2007). The Lojasiewicz Inequality for Nonsmooth Subanalytic Functions with Applications to Subgradient Dynamical Systems. Society for Industrial and Applied Mathematics.
- [6] J.F.Aujol, Ch.Dossal, A.Rondepierre (2022). Convergence Rates of the Heavy-Ball Method under the Lojasiewicz Property. Mathematical Programming.
- [7] J. Zhang, Tianxing He et Al (2020). Why Gradient Clipping Accelerates Training : A Theoretical Justification for Adaptivity. ICLR.

# RADIOMIC ANALYSIS FOR RELAPSE PREDICTION IN DIFFUSE LARGE B-CELL LYMPHOMA

**Khaoula CHAHDI**

**Direction : Olivier SAUT**

**Encadrement : Christèle ETCHEGARAY**

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

**L**arge B-cell lymphoma is a type of Non-Hodkin lymphoma which is a blood cancer that affects the lymphatic system where the body makes abnormal B-lymphocytes. This cancer is aggressive and progressive and characterized by a multifocal nature with intra and inter-lesion texture heterogeneity. As medical imaging is a useful tool for diagnosis in a non-invasive way and contains a huge amount of information not yet investigated, recent researches especially in oncology started focusing on the use of Radiomic features that are not visible to the naked eye and extracted from medical imaging with mathematical methods for the purposes of increasing their predictive potential. Since patients with Large B-cell lymphoma are at high risk of relapse, we aim in this study to set up a Machine Learning model based on Radiomics that can predict patient relapse and the relevant features, the thing that can enhance Radiomics-based decision for precision diagnosis and treatment.

## Introduction

In this work, we are going to explore the use of Radiomics to enhance the prediction of relapse for patients with Large B-Cell lymphoma.

## Materials and Methods

### Study population

In this study 80 patients diagnosed and treated in CHU of Bordeaux. 18F FDG PET/CT was performed in these patients as a component of baseline staging. The segmentation of these images was semi-automatic by thresholding and manually cleaned by physicians ensuring the registration of images from both modalities. The number and location of lesions vary widely from patient to patient.

### Features Extraction and Selection

Once the method of proceeding is established and the ROI is set, the binary mask and the image from one of the modalities are given to a function of Pyradiomics platform to extract the Radiomics. These characteristics are divided into several groups, histogram and shape and texture, we end up with more than 107 features compared with the number of patients we are left with the curse of dimensionality that may impact the model ability to learn. In this case the dimensionality reduction is necessary by removing pairwise correlated features and features that provide redundant information. (see Figure 40.1, 40.2 and table 40.1)

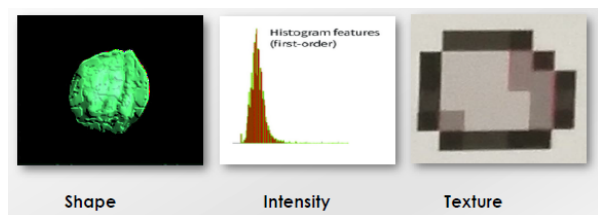


FIGURE 40.1 – Different groups of Radiomic features.

### Machine learning model

A statistical model fitted to predict the desired outcome of the study, in this case, patients relapse. The algorithm basically used in our methods is the logistic regression which is suitable for the low number of patients. The metrics used to evaluate the performance of our models are ROC-AUC curve, and the Brier score to measure their calibration. The model is trained on a set of data, then its performance is tested on another set not seen during the training phase, in our case we use 10-fold cross-validation to avoid the bias.

### Methods

We've proceeded with three different methods, the first one we used a mask of combined lesions that we considered as one lesion for each patient then we extracted the Radiomics based on this mask, and in the second one we considered only the biggest lesion of every patient. As this method does not take into account the spatial distribution of the lesions a Multi-Instance method was used to preserve this information.

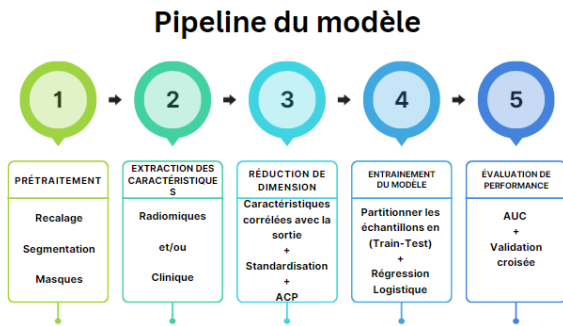


FIGURE 40.2 – A diagram summarizing the various steps.

## Results

### Clinical Model

Among the information provided by physicians we have clinical data such as sex, age, Dmax which is the maximum distance between lesions for each patient and TMTV (Total metabolic tumor volume) and TLG (total lesion glycolysis). We built a model based only on this data as explained previously. We used the AUC as a metric of evaluation in a 10-fold cross validation.

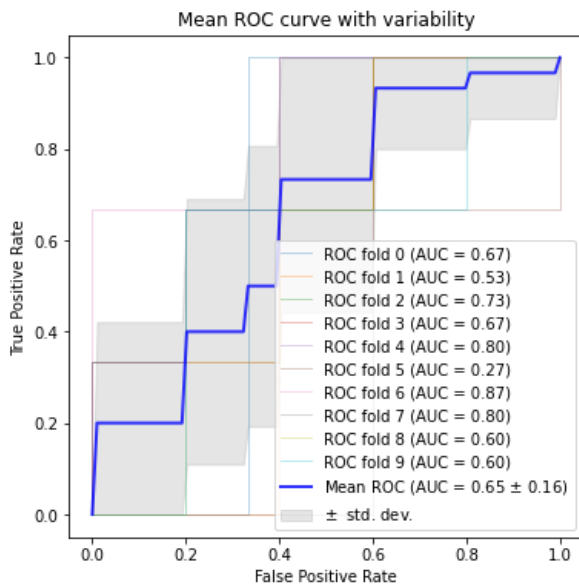


FIGURE 40.3 – ROC-AUC curve of clinical model.

### Multi-Instance Radiomics Model

In this method, we considered all the lesion for each patient we calculated the Radiomics for each lesion separately then we aggregated them with a MIL method. Same as the clinical method we used AUC as metric and 10-fold cross validation for the evaluation.

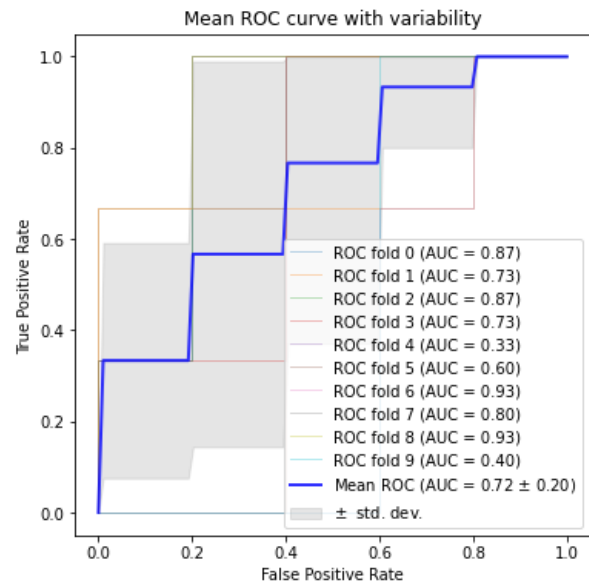


FIGURE 40.4 – ROC-AUC curve of the MIL model with Pet scan Radiomics

### Models results and Comparison

| Model     | AUC  | std  |
|-----------|------|------|
| Clinical  | 0.65 | 0.16 |
| Radiomics | 0.72 | 0.2  |

TABLE 40.1 – AUC table of the two models clinical and radiomics

We can see that comparing to clinical model, the Radiomics model with MIL improved significantly the prediction of the relaps. The relevant feature with the Pet scan for this model are shape features. There will be a comparison with Radiomics of the CT scans since it has better resolution and thus more details about texture.

## Bibliographie

- [1] Im HJ, Bradshaw T, Solaiyappan M, Cho SY. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography : Which One is Better? Nucl Med Mol Imaging. 2018 Feb;52(1) :5-15. doi : 10.1007/s13139-017-0493-6. Epub 2017 Sep 19. PMID : 29391907; PMCID : PMC5777960.
- [2] Wagner MW, Namdar K, Biswas A, Monah S, Khalvati F, Ertl-Wagner BB. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. Neuroradiology. 2021 Dec;63(12) :1957-1967. doi : 10.1007/s00234-021-02813-9. Epub 2021 Sep 18. PMID : 34537858; PMCID : PMC8449698.

# FINITE VOLUME METHODS FOR THE CELL-BY-CELL MODEL OF CARDIAC ELECTROPHYSIOLOGY

**Zeina CHEHADE**

**Direction : Yves COUDIERE**

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

The electrical excitation of the myocardium is a pivotal mechanism for sustaining proper cardiac function. Related diseases such as arrhythmias or atrial fibrillation have major impact on heart failure and infarction, contributing significantly to mortality rates in developed countries. Numerical simulation of cardiac excitation mechanisms helps understanding of disease onset and progression, offering valuable tool for improving diagnosis and therapies. In this context, as part of the European project MICROCARD, the objective of my thesis is to develop suitable Finite Volume schemes ensuring numerical efficiency in terms of convergence order, sparsity, and condition numbers. Moreover, these schemes should be suited for current and future HPC architectures in terms of parallelization and vectorization opportunities.

## Introduction

### Scientific Context and Motivation

The standard mathematical models for describing the electrical activity of cardiac tissue, such as the bidomain and monodomain models, were established by homogenizing the myocardium. While they are very successful in reproducing many effects and mechanisms, some aspects caused by heterogeneous microstructure of the myocardium cannot be captured. This includes heterogeneity on the cellular scale as present in fibrosis, spatial distribution of ion channels and gap junctions, or the shape and interconnection of myocytes. For that reason, detailed models representing the discrete cellular structure of the myocardium have recently gained attention. The most prominent example is the Extracellular-Membrane-Intracellular (EMI) model (see [1]), representing all myocytes as individual subdomains, and treats the nonlinear dynamics of ion channels only on the subdomain membranes.

Indeed, the challenge here is that the cell-by-cell model has unique aspects requiring non-standard discretization techniques. First, the computational domain is heterogeneous, with different unknowns being associated to subdomains of varying dimension, and only piecewisely continuous solutions. Second, the temporal dynamics are restricted to the cellular membrane, while a global quasistatic electric balance must be ensured. Hence, semi-discretization in space yields a system of differential-algebraic equations (DAE).

Many research studies have proposed various approaches to address the EMI problem, among them the Finite Element Method (FEM) and the Boundary Element Method (BEM). To our knowledge, there have been no works covering the finite volume approach for the microscopic model, so that the originality of our work lies in being the first step towards implementing the FVM for the EMI model.

### Micro-scale Model

The EMI model consists of  $N_C + 1$  piecewise electrostatic equations coupled through non-standard time dependent transmission conditions on the cell membranes (see [3]). Assuming  $N_C = 1$ , for  $i = 0, 1$ , and  $T > 0$ , the model reads

$$\begin{cases} -\nabla \cdot (\sigma_i \nabla u_i) = 0, & \text{in } \Omega_i, \\ I_m := c_m \partial_t v + f(v, w), & \text{on } \Sigma, \\ -\sigma_0 \nabla u_0 \cdot n_0 = \sigma_1 \nabla u_1 \cdot n_1 = -I_m, & \text{on } \Sigma, \end{cases} \quad (41.1)$$

where  $\sigma_i$  stands for the subdomain conductivities,  $c_m$  the membrane capacitance, the function  $f$  for the ionic current, the state variables  $v = u_1 - u_0$  and  $w$  stand for the transmembrane voltage and gating/concentration variables respectively. This system is completed by mixed Dirichlet and Neumann inhomogeneous boundary datum,  $g^D$  on  $\Gamma^D$  and  $g^N$  on  $\Gamma^N$  resp., with initial datum  $v_0$  and  $w_0$  on the membrane  $\Sigma$ . The main physical unknowns are the electrical potentials  $u_i$ .

### Objective

In my thesis, we aim at developing suitable finite volume schemes, which is needed to be numerically efficient in terms of convergence order, sparsity, condition numbers, as well as suited for current and future HPC architectures in terms of parallelization and vectorization opportunities.

### Methodology

For the well-posedness proofs, a comprehensive range of functional analysis tools was employed, with reference to a previous works (see [2]). Spatial discretization involved the classical cell-centered scheme, Two-Point Flux Approximation (TPFA) for admissible meshes, beside more flexible schemes as Hybrid Finite Volume

(HFV) and other FV classes. Regarding time integration, Implicit-Explicit Euler (IMEX), Backward Differentiation Formula (BDF) and Rush Larsen (see [4]) time-stepping methods were used. Mesh generation, while adhering to admissibility constraints for the TPFA scheme, used a variety of software tools, including Triangle, Gmsh, Mmg, PyDistmMesh python library ... All together by some tools providing by MICROCARD partners specializing in Mmg to address the obtuse mesh angles. The codes used, based on a data structure already exists in CARDIOLIB Inria Software to handle and read meshes.

## Results and Contributions

Results of the first part show the well-posedness of the micro-model with one myocyte, including an extension to the case of a sequence of cardiac cells, with more general assumptions on the domain.

The second part presents a comprehensive and quite technical analysis of the convergence for the TPFA scheme (well-posedness of the discrete system, discrete norms, errors estimates, ...) with different regularity assumptions (see [1] for  $C^2$  regularity). The formulation of the TPFA discrete flux on the membrane, is expressed as a linear combination of the cells unknowns and the interface voltages which is exclusive for the FVM. Numerical results were obtained to confirm the theoretically predicted convergence rates. Numerical tests was conducted using various time integration methods (BDF, IMEX and RL), and several manufactured solutions tailored to the domain geometry and multiple geometrical settings. The convergence can be deduced from the existence proof already established, with fewer regularity assumptions on the solution and the prescribed data. The FVM naturally yield a large sparse linear system similar to that of FEM, or a much smaller one exclusively for the membrane unknowns, as observed in BEM. Hence, FV formulation allows us to write the governing system as an evolution one only on the cell membranes, computing the Shure complement and iterating with

$$\left( C - B^T A^{-1} B + \frac{c_m}{\Delta t} D \right) v_\Sigma^n = \left( \frac{c_m}{\delta t} v_\Sigma^{n-1} - f(v_\Sigma^{n-1}) \right) D - B^T A^{-1} G^n \quad (41.2)$$

which reduce the system size, as well as the computational cost. Moreover, we investigated a comprehensive comparison in both sparse and dense versions, in terms of nonzero elements, numerical solutions (are similar in both versions up to numerical precision), discrete errors, and convergence rates. Moreover, we computed some illustrative electrical wave propagation.

A real challenge encountered was in constructing the admissible constrained mesh. Several software tools were employed to enhance the mesh quality under such constraint, resulting in the successful identification of admissible meshes with various geometrical settings for FV4'scheme (ECCOMAS abstract).

Ongoing joint work with MICROCARD partners focuses on benchmarking space discretisation for the EMI mo-

del, comparing FVM with BEM and FEM (Benchmark paper).

A coding contributions has been made to the CARDIOLIB Inria Software, involving the implementation of the aforementioned numerical schemes and mesh generation scripts.

## Ongoing and Outlooks

In practice, the scheme is intended to be used for simulating billions of cardiac cells, using HPC solutions developed within MICROCARD consortium. Indeed, the ODEs from the second Eq. of (1.1) are coupled with large sets of nonlinear equations that requires time-integration specific to cardiac models. The FVM simplify the implementation of this coupling, by introducing the voltage  $v$  as an explicit unknown. The extension of TPFA method to higher dimensions proves challenging due to the admissibility constraint. Consequently, in the third part we intend to explore more flexible FVM, such as HFV (SUSHI for example) and polynomial construction based method, which can be generalized for higher dimensions, more sophisticated geometrical settings and general meshes.

## Acknowledgment

This work is part of the MICROCARD project. This project has received funding from the European High-Performance Computing Joint Undertaking EuroHPC (JU) under the grant agreement No 955495.

## Bibliographie

- [1] Chehade, Zeina and Coudière, Yves. (2023). The Two-Point Finite Volume Scheme for the Microscopic Bidomain Model of Electrophysiology. 10.1007/978-3-031-40864-9-18.
- [2] Becue, P.E. : Modélisation et simulation de l'électrophysiologie cardiaque à l'échelle microscopique. Theses, Université de Bordeaux (2018). URL <https://theses.hal.science/tel-02019648>.
- [3] Tveito, A., Jæger, K.H., Kuchta, M., Mardal, K.A., Rognes, M.E. : A cell-based framework for numerical modeling of electrical conduction in cardiac tissue. Frontiers in Physics 5 (2017). DOI 10.3389/fphy.2017.00048.
- [4] M. E. Marsh, S. T. Ziaratgahi, R. J. Spiteri. The secrets to the success of the rush-larsen method and its generalizations. IEEE Transactions on Biomedical Engineering, 59, 2506-2515, 2012.



# TOWARD ROBUST LINEAR IMPLICIT SCHEMES FOR STEADY STATES SUPERSONIC FLOWS.

**Benoit COSSART**

**Direction : Raphael LOUBÈRE**  
**Encadrement : Jean-Philippe BRAEUNIG**

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

**I**mplicit time discretization in computational fluid dynamics to compute steady state solution of hypersonic flows was an intense field of researches in the 70's-80's. Since then, only few novel methods have arisen. It is suitable for computational efficiency to use implicit finite volume schemes to compute steady state for hyperbolic systems, because theoretically these schemes do not suffer from time step restrictions to guarantee stability, unlike the explicit schemes. Nonetheless, CFL like time restrictions are still required in practice, especially for stiff numerical test cases such as high Mach number stationary flow with bow shocks around obstacles [3]. Nowadays, the well-known Yee's method [1] is commonly used to solve CFD problems the implicit way. However, this method has no formal theoretical basis for systems of conservation laws [2]. Consequently, time step is driven by several ad-hoc parameters depending on the test case. The explicit CFL constraint is still required to capture the solution, after which it can be slowly relaxed. The purpose of this work is to study the mathematical environment of implicit finite volume schemes to enlight their weaknesses and bring out a more adequate linearization processes [4] presented firstly for the Roe scheme [5], but in a general framework of finite volume schemes. Numerical 2D results will assess this study, in particular hypersonic Mach flows for reentry vehicles will be presented.

## Analysis of multi-D Euler equation linearization procedure

**Recall about Finite Volumes.** In this section we briefly recall the design of explicit Finite Volume schemes in 2D, more specifically when used with hypersonic fluid flows modeled by Euler system of PDEs. In this work we focus on the special case of the Euler hydrodynamics system of conservation laws in  $d = 2$ . The system of PDEs expresses the conservation of mass, momentum and total energy, and writes as :

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbb{F}(\mathbf{U}) = \mathbf{0} \quad (42.1)$$

Here,  $\mathbf{U}$  is the vector of conservative variables. This vector is such that  $\mathbf{U} = (\rho, \rho \mathbf{v}, \rho E)^t \in \mathbb{R}^{d+2}$  where  $\rho$  is the mass density,  $\mathbf{v} = (u, v)^t$  the velocity vector and  $E$  the specific total energy. The physical flux is described by the  $(d+2) \times d$  tensor,

$$\mathbb{F}(\mathbf{U}) = \begin{pmatrix} \rho \mathbf{v}^t \\ \rho \mathbf{v} \otimes \mathbf{v} + p \mathbb{I}_d \\ \rho E \mathbf{v}^t + p \mathbf{v}^t \end{pmatrix}. \quad (42.2)$$

Using, FV formalism on unstructured mesh, (1.1) is written at cell  $c$  as

$$\frac{\mathbf{U}_c^{n+1} - \mathbf{U}_c^n}{\Delta t} + \sum_f S_f \Phi_f^{n+1} \cdot \mathbf{n}_{cf} = 0, \quad (42.3)$$

where  $S_f$ ,  $\Phi_f^{n+1}$  and  $\mathbf{n}_{cf}$  are respectively the face surface, the numerical flux at time  $n+1$  and face  $f$  and the outgoing normal vector of cell  $c$  at face  $f$ .

**Analysis of linearization procedure.** The equation (1.3) is linearized writing  $\Phi_f^{n+1}$  in time increment form :

$$\Phi_f^{n+1} \cdot \mathbf{n}_{cf} = \Phi_f^n \cdot \mathbf{n}_{cf} + \Phi_f^{n+1} \cdot \mathbf{n}_{cf} - \Phi_f^n \cdot \mathbf{n}_{cf}, \quad (42.4)$$

and using the classic [1] linearization procedure,  $\Delta \mathbf{F}_c \approx \mathbb{J}_c^n \Delta \mathbf{U}_c$  where  $\mathbb{J}_c^n$  is the jacobian of the system at time  $n$  and cell  $c$ . The linearization procedure yields to a linear system :

$$\mathbb{A} \Delta \mathbf{U} = \mathbf{B}, \quad (42.5)$$

where  $\mathbb{A}$  is a bloc matrix of dimension  $(d+2) \times (d+2)$ . To analyse such a system, we propose a von Neumann type analysis on the 1D system for each of its component  $\alpha$ . In order to proceed with the analysis, we assume that block matrices are frozen at time  $t^n$  values. Let us consider one single Fourier mode  $k$  for cell  $j$  at time  $t^n$  of the vectors :  $\hat{\mathbf{U}}_j^{n+1} = \theta \hat{\mathbf{U}}_j^n$ , and,  $\hat{\mathbf{U}}_{j\pm 1}^n = e^{\pm i k \Delta x} \hat{\mathbf{U}}_j^n$ . Notice that vector  $\hat{\mathbf{U}}_j^n$  has three components. With analogy to von Neumann analysis, the amplification factors  $\theta_\alpha$  should be defined for each component  $\alpha \in (1, 2, 3)$

of vector  $\hat{\mathbf{U}}_j^n$ , i.e  $\hat{\mathbf{U}}_j^{n+1} = \begin{pmatrix} \theta_1 \hat{\rho}_j^n \\ \theta_2 \hat{\rho} u_j^n \\ \theta_3 \hat{\rho} e_j^n \end{pmatrix}$ . Unfortunately

considering the vector form leads to an unexploitable situation. Contrarily, in the following, we consider each component  $\alpha$  separately. Then assuming that for any

component  $\alpha$  we have

$$\left(\mathbb{A}_j \hat{\mathbf{U}}_j^{n+1}\right)_\alpha = \left(\mathbb{A}_j \begin{pmatrix} \theta_1 \hat{\rho}_j^n \\ \theta_2 \hat{\rho} \hat{u}_j^n \\ \theta_3 \hat{\rho} \hat{e}_j^n \end{pmatrix}\right)_\alpha \simeq \theta_\alpha \left(\mathbb{A}_j \hat{\mathbf{U}}_j^n\right)_\alpha. \quad (42.6)$$

Then the amplification factors expression leads to a stability criterion extended to multi-D as

$$\sum_f S_f \frac{(\mathbb{A}_{cf} \mathbf{U}_c)_\alpha}{(\mathbf{U}_c)_\alpha} \geq \left| \sum_f S_f \frac{(\mathbb{A}_{df} \mathbf{U}_c)_\alpha}{(\mathbf{U}_c)_\alpha} \right|, \quad (42.7)$$

where  $\mathbb{A}_{cf}$  and  $\mathbb{A}_{df}$  are respectively the diagonal and extra-diagonal matrices of matrix  $\mathbb{A}$  from 1.5. We noticed that this criterion is related to diagonal dominance property as it is an extension of scalar diagonal dominance property to system.

To respect this stability criterion, we proposed to use a monotone linearization process [4]. This implicit method presents an interesting property, that is  $\mathbb{A}_{cf} = \mathbb{A}_{df}$ . Hence, the criterion from 1.7 is respected by construction.

## Numerical Result

**Mach 20 bow shock around cylinder.** We compare the classic implicit scheme [1] to the monotone linearized implicit scheme [4]. The classic implicit linearized scheme does not respect the stability criterion (1.7) while the monotone implicit scheme does. The comparison is done over a 2D case involving a Mach 20 stationary bow shock around an half cylinder. The mesh is unstructured and composed 6 971 triangles. The inflow boundary conditions correspond to Mach 20 shock conditions, same as the computational domain initial conditions which is filled with diatomic gas at shock velocity speed  $u^0 = u_s = M_s \sqrt{\gamma \frac{p}{\rho}}$ , so  $(\rho^0, u^0, v^0, p^0, \gamma) = (1, u_s = 23.66, 0, 1, 1.4)$ . The implicit schemes are ran using FV-Roe numerical scheme to capture the steady state solution. We tailor the stiffest CFL profile allowed to run each linearization process without failing. The results are given in figure 1.1 and 1.2.

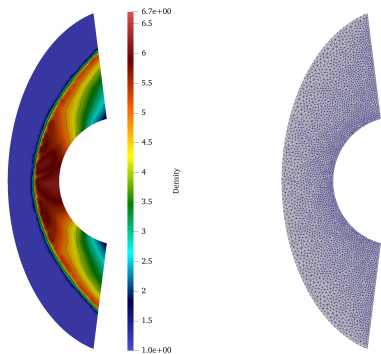


FIGURE 42.1 – 2D  $M = 20$  half cylinder density map and mesh — Numerical results using FV-Roe scheme.

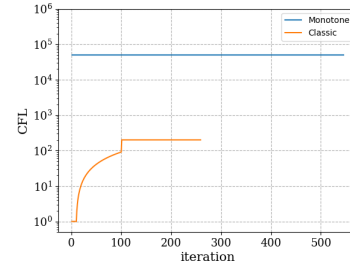


FIGURE 42.2 – CFL rampings

On figure 1.2, we see that we can only use the monotone linearization process with  $\text{CFL} > 1$ . Indeed, to ensure stability, the maximum CFL value for the classic linearization process is  $\text{CFL} = 1$ , which is the limit between explicit and implicit CFL values. For the monotone linearization process, we can run the computation using  $\text{CFL} = 50\,000$  from start and remains constant until the steady state is captured. Under the conditions of this test case, the classic linearization process is not capable to catch the  $M = 20$  shock wave, contrarily to the monotone linearization process, which can be used with very high implicit CFL values from start. To allow implicit values of CFL for the classic linearization process, we should define a CFL fine ramping.

Through preceding 2D test case, we see that the monotone implicit linearization procedure is more robust than the classic implicit linearization procedure. The monotone procedure respects by construction the stability criterion based on the analogy to diagonal dominance, while the classic linearization procedure does not. We assume the observed differences between those two implicit methods to be explained in the light of this stability criterion. We took this several test cases as illustrative of the relevance of stability criterion.

## Bibliographie

- [1] H.C. Yee (1989). A class of high-resolution explicit and implicit shock capturing methods. ICFD, Lecture Series, Von Karman Institute fo Fluid Dynamics.
- [2] A. Harten (1997), High Resolution Schemes for Hyperbolic Conservation Laws. Journal of Computational Physics, 135, 260–278.
- [3] Gisu Park, Sudhir L. Gai, Andrew J. Neely (2016), Base flow of circular cylinder at hypersonic speeds. AIAA Journal, 54.
- [4] H.C Yee and R.F Warming and A Harten (1985), Implicit total variation diminishing (TVD) schemes for steady-state calculations. Journal of Computational Physics, 3, 327–360.
- [5] P.L. Roe (1981), Approximate Riemann Solvers, Parameter vectors, and Difference schemes, Journal of Computational Physics, 43, 357–372.

# STATISTICAL LEARNING THROUGH THE SCOPE OF RANDOM MATRIX THEORY AND FREE PROBABILITY

Issa-Mbenard DABO

Direction : Jérémie BIGOT

Co-Direction : Camille MALE

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

**P**rediction is a classical problem in Statistical learning aiming to predict some target knowing covariates. In the era of Big Data, predicting models tend to have more and more parameters bringing both computational and mathematical issues. This yields the emergence of High Dimensional Statistics, a field studying efficient predicting models, methods and algorithms dealing with massive dimension of nowadays data. In this work, we propose to use Random Matrix Theory and Free Probability to study some regression models with non-identically distributed data in high dimension. Among these models, Neural network (NN) appeared to be efficient in various regression applications with moderate or high dimension. However, a certain number of open questions remain to be answered in NN learning. In fact, some misunderstood behaviors of NN modeling deserve investigations.

## Ridge regression with variance profile

We first focus on the following linear regression model

$$Y_n = X_n \theta_* + \varepsilon_n, \quad (43.1)$$

where  $X_n$  is  $n \times p$  matrix of random predictors,  $\varepsilon_n \in \mathbb{R}^n$  is a noise vector independent of  $X_n$  with  $\mathbb{E}[\varepsilon_n] = 0$  and  $\mathbb{E}[\varepsilon_n \varepsilon_n^t] = \sigma^2 I_n$ ,  $\theta_* \in \mathbb{R}^p$  is a vector of unknown parameters, and  $Y_n \in \mathbb{R}^n$  is the vector of observed responses. Classically, the predictors are assumed to be independent and identically distributed (iid) data, meaning that the rows of the matrix  $X_n$  are independent vectors sampled from the same probability distribution. In this paper, we propose to depart from this assumption by considering the setting where the rows of  $X_n$  are independent but non-identically distributed. To this end, we suppose that  $X_n$  is expressed in the following form

$$X_n = \Upsilon_n \circ Z_n,$$

where  $\circ$  denotes the Hadamard product between two matrices,  $Z_n = (Z_{ij})$  has iid centered entries with variance one, and  $\Upsilon_n = (\gamma_{ij}^{(n)})$  is a deterministic matrix. To simplify the notation, we shall sometimes write  $\gamma_{ij}^{(n)} = \gamma_{ij}$  and thus drop the (possibly) dependence of  $\gamma_{ij}$  on  $n$ . The matrix

$$\Gamma_n = (\gamma_{ij}^2) \in \mathbb{R}^{n \times p}$$

governs the variance of the entries of  $X_n$ , and it is called a *variance profile*.

We consider the high-dimensional context (with  $p$  growing to infinity at a rate proportional to  $n$ ) for which the least squares estimator is possibly not uniquely defined. Thus, we focus our analysis on the ridge regression estimator that is the minimizer of the following loss

function

$$\hat{\theta}_\lambda = \arg \min_{\theta \in \mathbb{R}^p} \frac{1}{n} \|Y_n - X_n \theta\|^2 + \lambda \|\theta\|^2,$$

for some regularization parameter  $\lambda > 0$ .

To study the statistical performances of the ridge regression estimator, we analyse its predictive risk defined as

$$\hat{r}_\lambda(X_n) = \mathbb{E}[(\tilde{Y} - \tilde{X}^t \hat{\theta}_\lambda)^2 | X_n], \quad (43.2)$$

where  $(\tilde{Y}, \tilde{X}) \in \mathbb{R} \times \mathbb{R}^p$  is independent from  $(Y_n, X_n)$  and satisfies

$$\tilde{Y} = \tilde{X}^t \theta_* + \tilde{\varepsilon},$$

with  $\mathbb{E}[\tilde{\varepsilon}] = 0$ ,  $\mathbb{E}[\tilde{\varepsilon}^2] = \sigma^2$  and  $\tilde{X}$ ,  $\tilde{\varepsilon}$  and  $\theta_*$  independent. In the above formula,  $\tilde{X} = \tilde{S}_p^{1/2} \tilde{Z}$  with  $\tilde{Z} \in \mathbb{R}^p$  a random vector with iid centered entries and variance one, and  $\tilde{S}_p = \mathbb{E}[\tilde{X} \tilde{X}^t] = \text{diag}(\tilde{\gamma}_1^2, \dots, \tilde{\gamma}_p^2)$  denotes the variance profile of  $\tilde{X}$ . Note that the risk  $\hat{r}_\lambda(X_n)$  is conditioned on the predictors  $X_n$ , and it is thus a random variable.

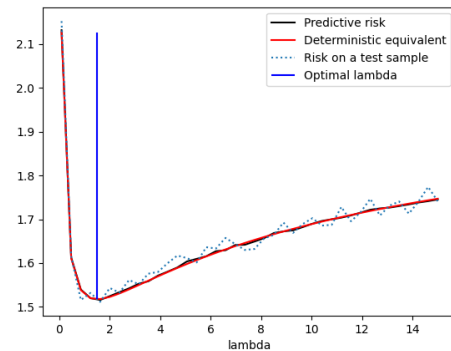


FIGURE 43.1 – Comparison of the predictive risk and its deterministic equivalent for various values of  $\lambda$ .

In this context, we are able to determine a deterministic equivalent of  $\hat{r}_\lambda(X_n)$  which can be easily computed numerically through the following theorem (see Figure 43.1).

**Theorem 43.0.1.** *Under some assumptions, there exists a deterministic matrix  $T_p(z)$  which allows us to provide the following deterministic equivalent of the deterministic risk*

$$r_\lambda(\Upsilon_n) = \sigma^2 + \frac{\sigma^2}{n} \text{Tr}[\tilde{S}_p T_p(-\lambda)] \quad (43.3) \\ + \lambda \left( \frac{\lambda \alpha^2}{p} - \frac{\sigma^2}{n} \right) \text{Tr}[\tilde{S}_p T'_p(-\lambda)],$$

in the sense that it satisfies

$$\lim_{n \rightarrow \infty, p/n \rightarrow c} |\hat{r}_\lambda(X_n) - r_\lambda(\Upsilon_n)| = 0, \text{ almost surely,}$$

where  $T'_p(z)$  denotes the derivative of  $T_p(z)$  with respect to  $z$ .

The matrix  $T_p(z)$  is defined in [3] which strongly inspired this result. A key argument to derive the computation of  $r_\lambda(\Upsilon_n)$  is to prove that the matrix  $T_p(z)$  is a relevant deterministic equivalent of the diagonal matrix  $\Delta[Q_p(z)]$  for an appropriate notion of asymptotic equivalence between matrices of growing size. This is done in [1] by using random matrix theory tools.

## Neural Networks and Free Probability

We are interested in the performance of high dimensional neural network. Thus we focus on the following model of single layer neural network

$$\mathbf{M}_n = \frac{1}{\sqrt{n}} \mathbf{g} \left( \frac{\mathbf{W}\mathbf{X}}{\sqrt{n}} \right),$$

where

- $\mathbf{g}$  is a smooth function such that  $\mathbb{E}[\mathbf{g}(\mathcal{N}(0, 1))] = 0$ .
- $\mathbf{W}$  is a  $m \times n$  matrix,  $\mathbf{X}$  is a  $n \times p$  matrix and their entries are centered i.i.d random variables.

The performance of the model in high dimension is related to the singular value distribution of  $M_n$  which is encoded by

$$\hat{\mu}_{M_n} = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k},$$

where  $\delta_x$  is the Dirac measure supported on  $\{x\}$  and  $\lambda_k$  is the  $k$ -th singular value of  $M_n$ . Since we are interested by the high dimensional framework, we study the behavior of  $\hat{\mu}_{M_n}$  when  $n, m$  and  $p$  tends to  $+\infty$ . According to Free Probability Theory, this behavior can be retrieve from the computations of the following moments

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathbb{E}[\text{Tr}((M_n M_n^*)^q)],$$

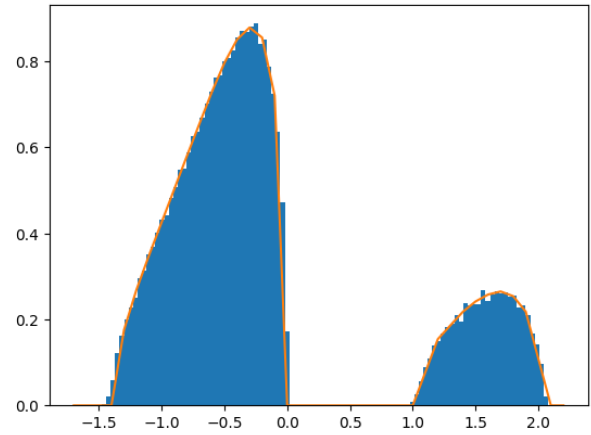
for  $q \in \mathbb{N}$ . We use these limits in order to extend a result from [2] by proving the following theorem which gives a random matrix model whose spectrum is similar to  $M_n$  spectrum.

**Theorem 43.0.2.** *Under some assumptions,  $M_n$  has the same limiting singular value distribution as the following matrix*

$$\sqrt{\theta_2(g)} \frac{\mathbf{W}\mathbf{X}}{\sqrt{np}} + \sqrt{\theta_1(g) - \theta_2(g)} \frac{\mathbf{Z}}{\sqrt{p}} + \theta_4(g) \frac{\mathbf{J}}{\sqrt{np}},$$

where  $\theta_1(g) = \mathbb{E}[g^2(\mathcal{N}(0, 1))]$ ,  $\theta_2(g) = \mathbb{E}[g'(\mathcal{N}(0, 1))]$ ,  $\theta_4(g) = \mathbb{E}[\frac{K}{6} g'''(\mathcal{N}(0, 1))]$ ,  $\mathbf{Z} \in \mathbb{R}^{m \times p}$  with i.i.d  $\mathcal{N}(0, 1)$  entries and  $\mathbf{J} \in \mathbb{R}^{m \times p}$  with entries equal to 1.

Thanks to this new random matrix model, we obtain a deterministic equivalent of the limiting singular value distribution of  $M_n$ . We are then able to compute numerically an accurate estimation of this singular value distribution (see Figure 43.2). Therefore we can evaluate the performance of the neural network in the high dimensional framework. We are currently willing to study a more general neural network model by endowing  $\mathbf{W}$  and  $\mathbf{X}$  with variance profiles. We have good hope that Theorem 43.0.2 can be extended in this context however this is not trivial.



**FIGURE 43.2** – Comparison of the histogram of singular values of  $M_n$  and its deterministic equivalent (orange line).

## Bibliographie

- [1] Bigot, J., Dabo, I., Male, C. (2024). High-dimensional analysis of ridge regression for non-identically distributed data with a variance profile, arXiv preprint arXiv :2403.20200.
- [2] Benigni, L., P    , S. (2021). Eigenvalue distribution of some nonlinear models of random matrices. Electronic Journal of Probability, 26 :1-37.
- [3] Hachem, W., Loubaton, P., and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. The Annals of Applied Probability, 17(3) :875 - 930.

# MDAE : MODIFIED DENOISING AUTOENCODER FOR MISSING DATA IMPUTATION

Mariette DUPUY

Direction : Marie CHAVENT<sup>1</sup>  
Co-Direction : Rémi DUBOIS<sup>2</sup>

1- INRIA Bordeaux Sud-Ouest, Talence, France

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

With the rapid increase in data collection, the presence of missing values is an ubiquitous challenge across various domains. One usually prefers missing data imputation which consist in filling missing entries with estimated values using the observed data. We propose a modified Denoising AutoEncoders, called mDAE, dedicated to missing data imputation.

## Introduction

With the rapid increase in data collection, the presence of missing values is an ubiquitous challenge across various domains. Data may be missing for several reasons as for instance it was never collected, records were lost or merging several datasets failed. It is then generally necessary to deal with this problem before performing machine learning methods on these data. Several options are available to address this issue including removing or recreating the missing values. One usually prefers missing data imputation which consist in filling missing entries with estimated values using the observed data. We propose a modified Denoising AutoEncoders, called mDAE, dedicated to missing data imputation. AutoEncoders (AE) [1] are artificial neural network used to learn efficient representation of unlabeled data (encodings) and a decoding function that recreates the input data from the encoded representation. Denoising AutoEncoders (DAE) where first proposed by [2] to recover, from noisy data, the original data without noise, by simply corrupting the inputs of a standard AE. Our mDAE is a DAE that one can apply on input data with missing data.

## Theoretical background

### The DAE

DAEs [2] are a variant of AEs developed to build a network trained to reconstruct a corrupted version of input data. A DAE is, therefore, an AE that receives corrupted input data. The corruption is done by changing the initial input  $\mathbf{x}$  to get a noisy version  $N(\mathbf{x})$ . This is done using a stochastic mapping  $N(\mathbf{x}) \sim q_D(N(\mathbf{x})|\mathbf{x})$  [2]. In our experiments, we consider a corrupting process named **masking noise**. During this process, a fixed proportion  $\mu$  of component is chosen at random and their value is forced to 0, while the others are left untouched.  $\mu$  is an hyper-parameter that can be optimized. This corrupting process is motivated by the goal of learning representations of the input that are robust to small

irrelevant changes in input. For [2], adding a denoising criterion helps to capture interesting structure in the input distribution. This leads to intermediate representations much better suited.

In the DAE, the corrupted input  $N(\mathbf{x})$  is mapped, as with the basic AE, to a hidden representation  $\mathbf{y} = f_\theta(N(\mathbf{x})) = s(\mathbf{W}N(\mathbf{x}) + \mathbf{b})$  from which we reconstruct a  $\mathbf{z} = g_{\theta'}(\mathbf{y}) = s(\mathbf{W}'\mathbf{y} + \mathbf{b}')$ . As before the parameters are trained to minimize the **reconstruction error**  $\|\mathbf{X} - \mathbf{Z}\|_F^2$ , i.e to have  $\mathbf{z}$  as close as possible to the incorrupted input  $\mathbf{x}$ . The Figure 44.1 represent a scheme of a DAE.

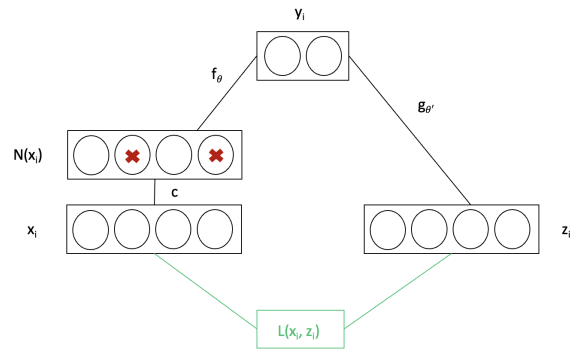


FIGURE 44.1 – Scheme of a DAE.  $c$  is the corrupting process of the input data  $\mathbf{x}$  : the masking noise. In this case,  $\mu = 0, 5$ . The red crosses represent the components forced to zero.  $\mathbf{y}$  is the latent space.  $\mathbf{z}$  is the output.  $L$  is the reconstruction error.

### The mDAE

With the description of a DAE, we can proceed to explain our method : the mDAE. More than just an algorithm, we aimed to present a turnkey method to use on numerical datasets containing missing values.

**Notations.** Let  $\mathbf{X}_{n \times p}$  be a matrix of centered (or centered-reduced) incomplete data, and  $\Omega \subset 1, \dots, n \times 1, \dots, p$  be the set of indices where the data is not missing. We define a matrix  $P_\Omega(\mathbf{X})$  of dimension  $n \times p$  as



follows :

$$P_{\Omega}(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{if } (i, j) \notin \Omega \end{cases}$$

which we will refer to as the projection of the matrix  $\mathbf{X}$  onto the available data. Similarly, we define the complementary projection  $P_{\Omega}^{\perp}(\mathbf{X})$  through  $P_{\Omega}^{\perp}(\mathbf{X}) + P_{\Omega}(\mathbf{X}) = \mathbf{X}$ .

The imputed data will then be :

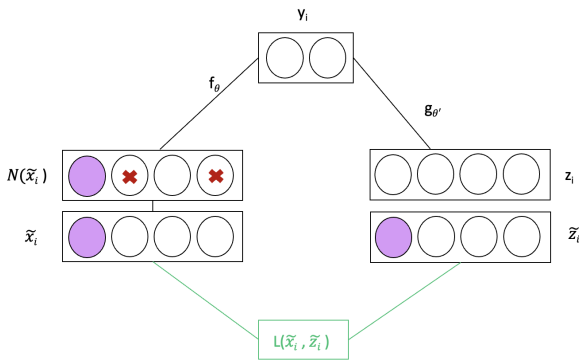
$$P_{\Omega}(\mathbf{X}) + P_{\Omega}^{\perp}(\mathbf{Z}),$$

where  $\mathbf{Z}$  is the matrix reconstructed by the network.

**The method.** The mDAE is an imputation method for numerical data with missing one (Figure 44.2). First step is the **standardization**. Each column is with the mean and the variance of the column itself regardless the presence of missing data. A neural network can not work with "NA's". The method propose a step of **pre-imputation**. The missing data of a same column are replaced by the mean of this column. As the data set is standardized, the mean equals 0. The missing data are replaced by 0's. At this stage the data set is noted  $P_{\Omega}(\mathbf{X})$ . It is on  $P_{\Omega}(\mathbf{X})$  that is applied the **DAE**. A proportion  $\mu$  of masking noise is declared among the indices of  $\Omega$ . Like in the former subsection, the encoder and the decoder optimize the parameters of the neural network to give an output  $\mathbf{z}$ . The specificity of mDAE is to modify  $\mathbf{z}$  by **forcing the component** that do not belong to  $\Omega$  to 0 :  $P_{\Omega}(\mathbf{Z})$ . The optimized **reconstruction error** is then :

$$\mathcal{L} = \|P_{\Omega}(\mathbf{X}) - P_{\Omega}(\mathbf{Z})\|_F^2$$

By modifying the output of the DAE, we allow the use of data set with missing data as input of the mDAE. The loss is not perturbed by the presence of missing data.



**FIGURE 44.2 – Scheme of a mDAE.** The violet dots represent the missing data equal to 0. The input data is  $P_{\Omega}(\mathbf{X})$ . For the corrupting process  $\mu = 0,5$  of masking noise is applied. The output  $\mathbf{z}$  is modified to get  $P_{\Omega}(\mathbf{Z})$ . The reconstruction error  $L$  is calculated between  $P_{\Omega}(\mathbf{X})$  and  $P_{\Omega}(\mathbf{Z})$

## Results

Our method is evaluated on UCI datasets<sup>11</sup>. UC Irvine Machine Learning Repository maintain 664 datasets as a service to the machine learning community. To compare mDAE to others baselines and evaluate the quality of imputation we use metrics. The widely used **Mean Absolute Error (MAE)**. Let  $m_0$  denote the total number of missing values artificially put in a dataset  $d$ . We define MAE as :

$$MAE_d = \frac{1}{m_0} \|P_{\Omega}^{\perp}(\mathbf{X}) - P_{\Omega}^{\perp}(\mathbf{Z})\|_F^2.$$

In the procedure of comparing imputation method we need to artificially add missing data in order to control the quantity, the type and to have the true value (Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR)). To evaluate and compare the mDAE with competing methods, we start with complete datasets and generate missing data. In the results we see that our method mDAE always outperforms the methods based on Generative Neural Networks and the methods of Optimal Transport. Surprisingly (never seen in other articles) methods of standard machine learning outperforms all the other methods on all the datasets.

## Bibliographie

- [1] Yoshua Bengio et al. (2009) Learning deep architectures for ai. Foundations and trends® in Machine Learning, (1) :1-127.
- [2] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. (2008) Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103.

11. <https://archive.ics.uci.edu/datasets>

# PARAMETRIZATION FREE STRUCTURE-TEXTURE DECOMPOSITION

Antoine GUENNEC

Direction : Jean-François AUJOL

Co-Direction : Yann TRAONMILIN

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

In this thesis, we study the problem of structure-texture decomposition, where we wish to decompose the image into two components. Tuning the parameters in the decomposition methods can be time consuming and disrupt the larger process the decomposition algorithm may be part of. We investigate different strategies to free ourselves from the tuning parameters that balance the structure and texture output. In a first part, we show that the tuning parameters in the low patch rank model can be automatically tuned by estimating the gradient sparsity of the structure component and estimating the patch-rank of the texture component. In a second part, we examine the mixed structure-texture model, where we have a unique regularization that takes both component into account.

## Structure texture decomposition

### Background

The problem of image structure-texture decomposition is stated as follows : given an image  $f \in \mathbb{R}^{n,m}$ , find a decomposition  $f = u + v$  such that  $u$  contains information on the general content in the image and  $v$  contains the textural information in the image. Mathematically, we ask that  $u$  is piecewise constant or smooth and  $v$  is a zero-mean, highly oscillating, with local patterns. The problem is notoriously ill posed as the associated system is under-determined. Hence, setting constraints upon  $u$  and  $v$  is necessary in order for them to be recovered.

The decomposition is useful in many image processing applications, for example :

1. Image/Video compression : the structure and texture components provide different visual information and belong to two low dimensional models. This allows for an efficient compression, without losing too much perceptual differences.
2. Material recognition : the texture layer provides important information regarding the material.
3. Medical imaging : the structure or texture may contain unnecessary details regarding the diagnosis.
4. Texture enhancement/removal : extracting the texture from the image allows to enhance/remove some details without interfering with the geometry of the image.

A popular way to frame image decomposition is by considering it as a low-dimensional model recovery problem. In this setting, we assume that  $f = u_0 + v_0$  with  $u_0 \in \Sigma_s$  and  $v_0 \in \Sigma_t$ , where  $\Sigma_s$  and  $\Sigma_t$  are two low-dimensional models/spaces. In this configuration, we aim to recover  $u_0$  and  $v_0$  via solving an optimization problem of the

form

$$\min_{(u,v)} \mu R_s(u) + \gamma R_t(v) \quad \text{subject to } f = u + v, \quad (45.1)$$

where  $R_s$  and  $R_t$  are regularization functions that aim to enforce  $u$  and  $v$  towards  $\Sigma_s$  and  $\Sigma_t$  respectively, and  $\mu, \gamma$  are the tuning parameters.

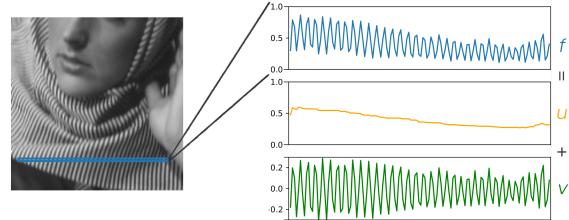


FIGURE 45.1 – Illustration of the structure-texture decomposition

### Motivations

Currently, structure-decomposition models suffer a range of shortcomings. Most notoriously, in order to obtain a satisfactory decomposition a lot of tweaking of the many different tuning parameters has to be done. This is especially important as the computations may be expensive when dealing with large amounts of data (e.g. biomedical images) or if the decomposition is but a step in a larger procedure.

### Adaptive parameter selection for gradient-sparse + low patch rank recovery

In this work, we are interested in gradient sparse + low patch-rank signal recovery for image structure-texture decomposition. Recent theoretical results on sparse + low rank recovery [2, 3] have shown that a sparse matrix  $\bar{u} \in \mathbb{R}^{n,m}$  such that  $\|\bar{u}\|_0 \leq s$  and a low rank matrix  $\bar{v} \in \mathbb{R}^{n \times m}$  such that  $\text{rank}(\bar{v}) \leq r$  can be recovered

from a linear measurement  $b = \mathcal{A}(\bar{u} + \bar{v})$  under certain conditions via the L1-nuclear norm minimization scheme

$$\min_{\substack{u,v \\ b=\mathcal{A}(u+v)}} \mu \|u\|_1 + \gamma \|v\|_* \quad (45.2)$$

when  $\frac{\mu}{\gamma} = c\sqrt{\frac{s}{r}}$ , where  $c$  is a specified constant.

Coincidentally, this result also holds in the context of the low patch-rank (LPR) interpretation of texture [1], where it is assumed that the set of patches of textures is of low patch-rank and the structure is gradient sparse. Indeed, numerical experiments [4] show that for a range of values  $c \in (0.5, 0.75)$ , a structure-texture couple  $(u_0, v_0)$  with a gradient sparsity  $s$  and a patch-rank  $r$  can be recovered from  $f = u_0 + v_0$ , in the same way to the sparse+low rank case, via the optimization problem

$$\min_{\substack{u,v \\ f=u+v}} \mu \|\nabla u\|_1 + \gamma \|\mathcal{P}v\|_*, \quad (45.3)$$

where  $\mathcal{P}$  is the patch operator.

By estimating the gradient sparsity of  $u$  and the patch-rank of  $v$  in the iteration of the minimization scheme of (45.3), we are able to estimate the optimal parameters  $(\mu, \gamma)$  and thus implement a near parameter-free implementation of the LPR model. More importantly, this strategy makes it possible to implement a localized version of the LPR model.



FIGURE 45.2 – Example of a decomposition obtained with the localized LPR model, with no additional tuning ( $c = 0.65$ ). From left to right : original image  $f$ , structure component  $u$ , the texture component  $v$

### Mixed structure-texture regularization

In a second part of the thesis, we investigate the mixed structure-texture model. In most decomposition models up to this point, the resulting decomposition was achieved via the minimization of the sum of two regularization (one for the structure component and one for the texture component), as in (45.1). However, this approach forces the introduction of tuning parameters in order to balance the output structure/texture. Furthermore, it considers the structure and texture as part of two very separate models, disregarding that in natural images the two components are actively linked by a shared support.

In the mixed structure-texture model, we explore the possibility of recovering the decomposition via the minimization of a single regularization which takes both cartoon and texture as input,

$$\min_{x=(u,v)^T} R(x) \quad \text{subject to } f = \mathcal{A}x, \quad (45.4)$$

where  $\mathcal{A} = (Id \ Id)$  is the composition operator. To construct such regularization, we adopt the gradient step denoiser from [5], where

- $R(x) = \frac{1}{2} \|x - N(x)\|$  is the regularization,
- $N : \mathbb{R}^{n,m} \rightarrow \mathbb{R}^{n,m}$  is parametrized by a neural network,
- $D(x) = (Id - \nabla R)(x)$  is the associated denoiser.

We learn the regularization by training  $D$  to denoise structure-texture couples, using a dataset comprised of randomly generated synthetic structure-texture couples that are created on-the-fly during the training process. Once the regularization trained, we may decompose the image via a forward-backward iterative scheme

$$x_{k+1} = \mathcal{P}_{C_f}(x_k - \lambda \nabla R(x_k)), \quad (45.5)$$

where  $\mathcal{P}_{C_f}$  is the orthogonal projection onto the convex set  $C_f = \{(u, v) \mid f = u + v\}$ .

### Bibliographie

- [1] H. Schaeffer, S. Osher. A low patch-rank interpretation of texture, SIAM Journal on Imaging Sciences, 2013.
- [2] J. Tanner, S. Vary. Compressed sensing of low-rank plus sparse matrices, Applied and Computational Harmonic Analysis, 2023.
- [3] Y. Traonmilin, R. Gribonval, S. Vaiteer. theory of optimal convex regularization for low-dimensional recovery, arXiv preprint, 2022.
- [4] Guennec, A. ; Aujol, J. ; Traonmilin, Y. Adaptive Parameter Selection for Gradient-Sparse + Low Patch-Rank Recovery : Application to Image Decomposition, HAL preprint, 2024.
- [5] S. Hurault, A. Leclaire, N. Papadakis. Gradient step denoiser for convergent plug-and-play, International Conference on Learning Representations, 2021.

# L'INFLUENCE DE LA STRUCTURE DU GROUPE DE GALOIS SUR LA RÉPARTITION DES AUTOMORPHISMES DE FROBENIUS

Mounir HAYANI

Direction : Florent JOUVE

Co-Direction : Daniel FIORILLI

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

**O**n étudie la répartition des automorphismes de Frobenius dans des extensions de corps de nombres. Étant donné une extension Galoisienne de corps de nombres  $L/K$ , des disparités inattendues apparaissent dès que le groupe de Galois  $G := \text{Gal}(L/K)$  satisfait des conditions provenant de la théorie des groupes.

## Introduction aux biais de Tchebychev

En 1853, Tchebychev cherchait à effectuer une comparaison entre les nombres premiers  $p \equiv 1 \pmod{4}$  et les nombres premiers  $p \equiv 3 \pmod{4}$ . Dans une lettre envoyée à Fuss [1], Tchebychev a noté que pour un  $x \geq 2$  donné, la quantité  $\pi(x; 4, 3)$ , décrivant le nombre des premiers  $p \leq x$  vérifiant  $p \equiv 3 \pmod{4}$  semble être plus grande que  $\pi(x; 4, 1)$  (le nombre des premiers  $p \leq x$  tels que  $p \equiv 1 \pmod{4}$ ). Depuis, cette disparité a été désignée sous le nom de biais de Tchebychev. Pour tout  $q \geq 3$ , Rubinstein et Sarnak ont étudié dans [5] la question générale du biais de Tchebychev entre les classes inversibles modulo  $q$ . En notant pour tout  $x \geq 2$

$$\pi(x; q, a) = \#\{p \leq x : p \equiv a \pmod{q}\},$$

Rubinstein et Sarnak ont prouvé, Sous l'hypothèse de Riemann généralisée, pour  $q \geq 3$ , et  $a \neq b$  deux classes inversibles modulo  $q$ , l'existence de la densité logarithmique  $\delta(q; a, b)$  définie par  $\lim_{X \rightarrow +\infty} \frac{1}{\log(X)} \int_2^X \mathbb{1}_{\pi(x; q, a) > \pi(x; q, b)}(t) \frac{dt}{t}$ . En supposant également une hypothèse d'indépendance linéaire sur les zéros non triviaux des fonctions  $L$  de Dirichlet, ils confirment l'observation originale de Tchebychev de l'existence d'un biais en faveur de  $a$  (c'est-à-dire  $\delta(q; a, b) > \frac{1}{2}$ ) si et seulement si  $a$  n'est pas un résidu quadratique mod  $q$  et  $b$  est un résidu quadratique mod  $q$ .

## Biais de Tchebyshev dans les corps de nombres

Suivant les idées de Rubinstein et Sarnak, Ng [4] a étudié le biais de Tchebychev dans le contexte des corps de nombres. Soit  $L/K$  une extension galoisienne de corps de nombres de groupe de Galois  $G$ . Considérons  $C \subset G$  une classe de conjugaison dans  $G$ . Définissons la fonction de comptage des idéaux premiers pour tout

$x \geq 2$  :

$$\pi(x; L/K; C) := \sum_{\substack{\mathfrak{p} \leq O_K \\ N\mathfrak{p} \leq x}} \mathbb{1}_C(\varphi_{\mathfrak{p}}),$$

où  $\mathbb{1}_C$  est la fonction indicatrice de  $C$ . Ici la somme porte sur les idéaux premiers non nuls  $\mathfrak{p}$  de  $O_K$  et  $\varphi_{\mathfrak{p}}$  désigne la classe de conjugaison de Frobenius correspondante.

En notant  $\mathcal{P}_{C_1, C_2}$  l'ensemble

$$\left\{ x \geq 2 : \frac{\pi(x; L/K; C_1)}{|C_1|} > \frac{\pi(x; L/K; C_2)}{|C_2|} \right\},$$

la densité logarithmique de  $\mathcal{P}_{C_1, C_2}$  (lorsqu'elle existe) est la généralisation aux extensions galoisiennes de corps de nombres de  $\delta(q; a, b)$ , c'est-à-dire

$$\delta(\mathcal{P}_{C_1, C_2}) := \lim_{X \rightarrow +\infty} \frac{1}{\log X} \int_2^X \mathbb{1}_{\mathcal{P}_{C_1, C_2}}(x) \frac{dx}{x}.$$

Dans sa thèse [4], Ng a étudié les densités de type  $\delta(\mathcal{P}_{C_1, C_2})$ . En supposant la conjecture d'holomorphic d'Artin, GRH, ainsi qu'une hypothèse d'indépendance linéaire sur l'ensemble des zéros des fonctions  $L$  d'Artin, Ng a prouvé dans [4] l'existence de  $\delta(\mathcal{P}_{C_1, C_2})$  et a calculé sa valeur dans plusieurs cas explicites. En 2020, Fiorilli et Jouve ([2]) montrent, en ne supposant aucune hypothèse, qu'il existe une infinité d'extensions galoisiennes de corps de nombres présentant un biais de Tchebychev extrême ( $\delta(\mathcal{P}_{C_1, C_2}) = 1$ ).

## Prolongement des travaux de Fiorilli et Jouve

Dans ma thèse j'étudie les possibles généralisations du travail de Fiorilli et Jouve [2]. Leur résultat principale étant la possibilité de construire des extensions galoisiennes de corps de nombre  $L/K$  ayant un biais extrême sous des conditions de théorie de groupes sur le groupe de Galois  $G := \text{Gal}(L/K)$ . Plus précisément, ils prouvent, pour une tour  $L/K/\mathbb{Q}$  de corps

de nombres telle que  $L/\mathbb{Q}$  est galoisienne de groupe de Galois  $G^+$ , que si deux classes de conjugaison  $C_1$  et  $C_2$  dans  $G = \text{Gal}(L/K)$  ne possèdent pas le même nombre de racines carrées dans  $G$  mais sont contenues dans une classe de conjugaison commune  $C^+$  de  $G^+$ , alors  $L/K$  présente un biais de Tchebychev extrême par rapport à  $C_1$  et  $C_2$ . Mon travail [3] consiste à alléger ces hypothèses sur  $G$  et  $C_1$  et  $C_2$ . Une conséquence de mes résultats est que pour tout groupe  $G$  de cardinal  $|G| \geq 7$  il existe une extension galoisienne, que je construis explicitement, de groupe de Galois  $G$ , dans laquelle des cas extrême pour le biais de Chebyshev apparaissent. Mon travail actuel consiste à généraliser ces résultats lorsqu'on regarde un nombre quelconque de classes de conjugaisons au lieu de deux classes  $C_1$  et  $C_2$ .

## Bibliographie

- [1] Chebyshev, P., (1853), Lettre de M. le Professeur Tchébychev à M. Fuss sur un nouveau théorème relatif aux nombres premiers contenus dans les formes  $4n + 1$  et  $4n + 3$ , Bull. Classe Phys. Acad. Imp. Sci. St. Petersburg, 208.
- [2] Fiorilli, D., Jouve, F., (2022), Unconditional Chebyshev biases in number fields., J. Éc. polytech. Math. ,671–679
- [3] Hayani, M., (2024), On the influence of the Galois group structure on the Chebyshev bias in number fields, prépublication, <https://inria.hal.science/hal-04538383>
- [4] Ng, N., (2000), Limiting distributions and zeros of Artin L-functions, thèse de doctorat, University of British Columbia.
- [5] Rubinstein, M., Sarnak, P., (1994), Chebyshev's bias, Experiment. Math.,173–197.



# COMPUTER VISION FOR ELECTROPORATION-BASED ABLATION OF DEEP-SEATED TUMOUR

Eloïse INACIO

Direction : Baudouin DENIS DE SENNEVILLE

Co-Direction : Clair POIGNARD

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

This work aims to provide efficient computer vision tools to facilitate the electroporation-based ablation of deep-seated tumours. Intraoperatively, this procedure requires the localization of the electrodes used to deliver the electric pulses, and the registration of the cone beam computed tomographies captured before -where we easily segment the tumour and the organ of interest- and after the needle insertion. In this work, we associate deep learning with more traditional methods, such as the Hough transform or variational problems, to extract useful information while interventional radiologists perform the procedure. This involves fast programs, enabling a precise estimation of the effective treatment area, to guide the therapy in real-time.

## Introduction

Irreversible electroporation (IRE) ablation is a non-thermal ablation technique using electromagnetic fields to create irreparable pores in cells, leading to cell death. It is non-thermal and can thus be used for deep-seated tumours, near vital structures. However, it requires careful planning and evaluation, due to the presence of electric pulses.

This work provides computer vision tools to facilitate IRE evaluation in the operation room. As seen in Figure 47.1, there are several needs :

- Tumor and organ segmentation, i.e tumour ROI extraction (currently done semi-automatically)
- Needle localisation
- Computed Cone Beam Tomography (CBCT) registration i.e. ROI propagation

Here, we propose solutions to the needle localisation and image registration problems by combining deep learning approaches and more traditional methods.

## Needle localisation

We propose a coarse-to-fine approach to locating the needle tip delivering the electric pulses for IRE.

### Coarse segmentation

Firstly, we segment the needle on the CBCT scan using deep learning, that is, we attribute a class, either needle or background, to every pixel of the 3D image. A U-Net is adapted to the data and task at hand. Zero-padded 3D convolutions and non-overlapping patches are used to avoid unnecessary cropping steps and accommodate the CBCT scan dimensions. Also, twice as many decoder features are concatenated with the encoder features, given that extremely thin objects are of interest (0.9-1.3mm).

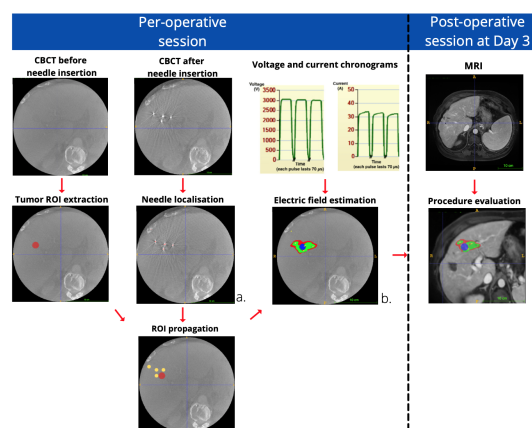


FIGURE 47.1 – Clinical and numerical workflow for IRE ablation of hepatocellular carcinoma (HCC) [1].

The training strategy is thought through to counter the class imbalance in the data. The combo loss, associating Dice loss and binary cross-entropy, is chosen. Moreover, the training data set consists of patches selected according to the information they carry, i.e. needle or background, to increase the representation of the needle class artificially. An equilibrium is reached for 50% needle patch.

Also, the inference threshold is increased in favour of the needle class to take into account that, despite the patch selection strategy the network had more opportunity to learn the background class.

### Fine localisation

Lastly, the needle tip is precisely localised using the segmentation mask as input to a Hough transform with a voting procedure. The Hough transform counts every point belonging to every parametrised shape in the image.

The needles are approximated as lines and parametrised using Robert's optimal line representation. After

selecting  $k$  times the number of needles, a vote takes place according to the number of points belonging to the line, the distance between extremities (i.e. needle length), the mean distance between the points and their projections on the line, and the distance between the extremity and the centre of the image, since we are interested in the needle tip, close to the tumour.

## Results

Compared to the previously used thresholding method with the same Hough algorithm, the machine learning-based coarse segmentation allows more robust, more precise and more stable localisation, thus improving the electromagnetic field estimation. The main control parameters appear to be the loss function, the needle-to-background ratio and the inference threshold in the training set.

## CBCT registration

Image registration consists of finding a transformation  $T$ , taking a moving image  $J$  to a fix image  $I$ . It may be formulated as a variational problem where we aim to minimise an energy, generally composed of a data fidelity term to have the registered image resemble the fix image, and a regularisation term to make the transformation more plausible. In this work, we use a gradient-based data fidelity term [2] to remain applicable to multi-modal registration tasks, and a diffusion regularisation.

## Boundary conditions

Medical imaging often presents partial fields of view due to capture limitations and/or dose exposure. In these cases, suitable boundary conditions are crucial to controlling error propagation towards the centre. However, both most commonly used boundary conditions, namely homogeneous Dirichlet and homogeneous Neumann, are too restrictive and do not allow prior information to be taken into account.

Here, we propose adaptive boundary conditions using flow fields detected via a motion inverse consistency map  $g$ . It takes the form of the following Fourier-Robin boundary condition :

$$\mathbf{A}_s \nabla T_s \cdot \mathbf{n} + (1 - \mathbf{A}_s) T_s = \gamma g_s, \text{ on } \partial\Omega \quad (47.1)$$

for  $s \in x, y, z$ ,  $\gamma \in \mathbb{R}$  and  $\mathbf{A}_s : g_s \mapsto [0, 1]$ .

## Results

The proposed boundary conditions were tested on two different tasks : a mono-modal lung CBCT registration task and a multi-modal abdominal CBCT to MRI task. In both cases, the novel boundary condition successfully hindered the propagation of errors from the boundary to the centre and produced equivalent or better results than both homogeneous Neumann and homogeneous Dirichlet boundary conditions.

The hyper-parametrisation was successfully automated by minimising the energy through a grid search.

This framework, being independent of the registration algorithm, can be adapted to other registration solutions straightforwardly.

## Conclusion and perspectives

In this thesis, two novel computer vision tools are developed to improve the precision of electromagnetic field estimation, central to determining the efficacy of IRE ablation.

The addition of deep learning to the Hough transform with voting procedure allowed a more accurate localisation of the needle tip from where the electric pulses are delivered. The proposed learning strategy is applicable to all sorts of skewed data sets.

The careful consideration of boundary conditions in variational methods for registration led to better motion estimation, including in the regions of interest closer to the image centre. In future work, we will evaluate the effect of such adaptive boundary conditions on the electromagnetic field estimation in the case of IRE ablation. Currently, the evaluation procedure, including the needle localisation, the CBCT registration and the electromagnetic field estimation, takes under 2 minutes. The proposed solution is thus applicable in clinical settings.

## Bibliographie

- [1] E Inacio, et al., Jun 2022, IEEE - MELECON 2022 - 21st Mediterranean Electrotechnical Conference. (10.1109/MELECON53508.2022.9842866). (hal-03862360)
- [2] B Denis de Senneville et al., 2016 Phys. Med. Biol. 61 7377

# MODELLING OF AERO-ELASTICITY IN LARGE TRANSFORMATIONS BY A PARTITIONED COUPLING : APPLICATION TO LARGE WIND TURBINES.

**Caroline LE GUERN**

**Direction : Michel BERGMANN**

**Encadrement : Frederic BLONDEL, Nicolas GUY, Jean-Lou PFISTER**

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

---

**W**ind turbine developments have led to more powerful offshore rotor systems, creating complex aero-hydro-servo-elastic behaviors that require numerical analysis. Multi-physical problems can be modelled in several ways : through a "monolithic" approach that combines each domain, or through "partitioned" methods where the equations of each physics as well as their time integration are separated. Among the latter, various coupling algorithms exist such as the "CSS" (Conventional Serial Staggered) which imposes a common time step for coupling, or "sub-cycling" where the timesteps of the different physics can be separated. Different coupling time integration schemes are to be compared in the context of fluid/structure interactions on a large-scale wind turbine, where gain in computational time, stability and accuracy of the different methods will be discussed.

---

## Introduction

Offshore wind turbines, evolving rapidly, face new design challenges such as increasing rotor sizes with blades over 100 meters long, requiring consideration of a double constraint linking mass to structural stiffness. Designing these turbines necessitates a balance between blade mass and structural flexibility, making the accurate capture of structural dynamics under various conditions crucial. Numerical modeling, particularly for aeroelastic behaviors that account for interactions between the turbine's structure and aerodynamics, is essential. The aero-servo-elastic modelling tool used throughout this work is DeepLines Wind [4]. The structural aspect of the wind turbine, controlled by the ROSCO [3] controller, is modelled using a finite element method, where the tower and blades are modelled through 3D non-linear geometrically exact beam elements with large transformations for the latter. The aerodynamic side of the problem can be modelled through the Blade Element Momentum (BEM) method which is one of the most commonly used in wind turbine aerodynamic simulations due to its low computational cost. BEM is a well suited design tool, but the method's theory makes strong hypothesis such as constraining loads evaluation in a rotor plane which can be violated for large offshore (floating) wind turbines with highly flexible blades. In the present work, a less assumption-heavy code using an inviscid free vortex wake (FVW) method, CASTOR [1], is used for aerodynamic loads computations. While higher fidelity is ensured by the vortex method, its use in aeroelastic simulations makes overall computational costs higher, rendering it impractical as a viable design tool. In the framework of this thesis, the aim is to demonstrate how implementing alternative par-

tioned coupling techniques allows for computational time reduction when using the FVW method for aeroelastic simulation purposes, while maintaining sufficient accuracy and stability.

## Methodology

### *Free vortex wake code*

CASTOR is a GPU accelerated solver which uses an inviscid 3D FVW method based on the lifting-line theory. Solving Euler equations, which in this case are written in vorticity form, is achieved using a Lagrangian method, in which Lagrangian markers delimit sets of vortex filaments used to model the vorticity field behind the rotor 48.1 (more detail on the method is given in [1]). The high computational costs of the FVW come mainly from the coupling of such method to a servo-elastic solver which can impose small time steps in comparison to a stand alone vortex code. In such case, each coupling time step provokes the generation of an extra row of vortex filaments behind the rotor, making the wake more filament dense as the coupling time step gets smaller. To remedy this, several approaches can be considered such as filament merging methods, process that was used in [1], or partitioned time integration coupling schemes where different temporal discretisations are allowed. The latter is the retained technique for this work.

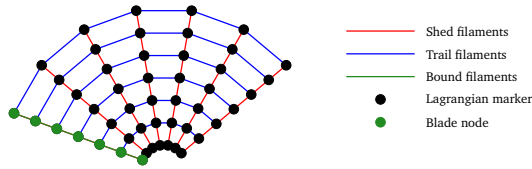


FIGURE 48.1 – Schematic representation of a rotating blade and the associated wake filament discretisation.

### Multi-physics coupling schemes

The CSS scheme [2] is one of the most commonly used amongst partitioned methods as it is one of the simplest configurations for sub-domain time integration separation. The different steps of such algorithm are depicted in 48.2. In order to go from time steps  $n$  to  $n + 1$ , four main steps are carried out :

- 1) structural displacements  $u^n$  and velocities  $\dot{u}^n$  computed at time  $t_n$  are transferred from the structural to the aerodynamic solver.
- 2) time integration of the fluid problem is conducted and aerodynamic forces  $F_f^{n+1}$  induced from structural displacements are obtained at time  $t_{n+1}$
- 3) aerodynamic forces are transferred to the structural solver  $F_s^{n+1}$
- 4) time integration of the structural problem is carried out to obtain structural displacements  $u^{n+1}$  induced from aerodynamic forces at time  $t_{n+1}$

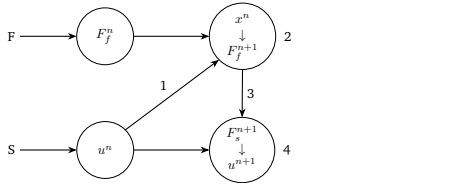


FIGURE 48.2 – Conventional Serial Staggered scheme between  $F$  a fluid solver and  $S$  a structural solver.

In 48.3 a revisited subcycling algorithm which was implemented in the DeepLines Wind tool is presented. At each coupling time step that corresponds to the aerodynamic's resolution rate, or  $\Delta t_f$ , the solver using the vortex method will directly communicate aerodynamic forces to the structural code. In the intermediate stages, which correspond to the structural solver's characteristic time ( $\Delta t_s$ ), a meta model library is itself in communication with DeepLines. The meta model estimates the state of the fluid in these intermediate stages by extrapolating computed aerodynamic forces from previous steps. Different types of regression techniques considered and implemented in this work were constant, linear and polynomial extrapolations but more sophisticated methods could also be implemented in this framework.

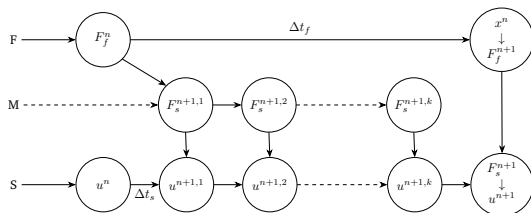


FIGURE 48.3 – Subcycling scheme with meta model.

## Results

In 48.4, a 200 s aeroelastic simulation, coupling DeepLines to CASTOR and modelling the IEA 15MW wind turbine, was run using various coupling algorithms. The considered reference case uses a CSS scheme, for which the coupling time step is set to  $\Delta t_c = 0.01$  s. A subcycling algorithm with a growing number of regression iterations using constant forecasting was implemented for comparison. The obtained sectional aerodynamic forces between the CSS and the subcycling methods show that as sub-iterations are added to the coupling algorithm, a slight difference in amplitude as well as phase offset due to constant meta model predictions is observed in comparison to the reference case. This is observed until a subcycled structure with 7 iterations, thus allowing the aerodynamics' integration time step to be set at  $\Delta t_f = 0.08$  s, above which the coupling algorithm diverges. Table 48.1 shows the total CPU times of the test case using the different coupling algorithms. By adding sub-iterations to the coupling scheme and thus allowing the fluid problem to be solved at a slower rate, total CPU time is drastically reduced with up to more than 250 times faster computing between a 7-iteration subcycled scheme and the CSS method. This is due to the reduction of wake filament emissions from the vortex method and the direct meta model use for aerodynamic data communication.

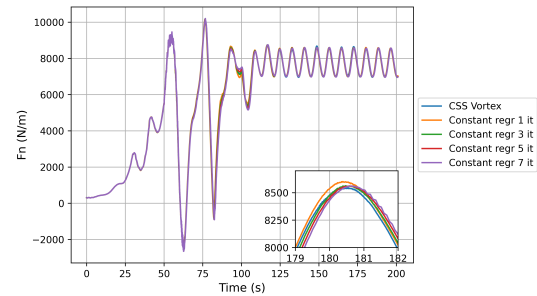


FIGURE 48.4 – Sectional aerodynamic forces comparison for CSS method and subcycling scheme using constant regression at radius location  $r = 70$  m.

| Method         | CSS    | 1 sub-it | 4 sub-it | 7 sub-it |
|----------------|--------|----------|----------|----------|
| Total CPU time | 10j16h | 35h06min | 2h55min  | 57min    |

TABLE 48.1 – Total CPU times for CSS and subcycled schemes.

## Bibliographie

- [1] Blondel, F., Joulin, P. A. and Le Guern, C. (2024). *TORQUE Conference (submitted)*
- [2] Heinz, Joachim, C., Sørensen, N. and Zahle, F. (2013). *PhD thesis, Technical University of Denmark*
- [3] Abbas, N. J. and Zalkind, D. S. and Pao, L. and Wright, A. (2022). *Wind Energy Science*
- [4] Le Cunff, C., Heurtier, J.M., Piriou, L., Berhault, C., Perdrizet, T., Teixeira, D., Ferrer, G. and Gillo-teaux, J.C. (2013). *OMAE*



# ARTIFICIAL INTELLIGENCE-ASSISTED TOOL FOR HEP-2 ANTINUCLEAR ANTIBODY PATTERNS' RECOGNITION

**Guillaume MARTINROCHE**

**Direction : Olivier SAUT<sup>1,2</sup>**

1. INRIA Bordeaux Sud-Ouest, Talence, France

2. Institut de Mathématiques de Bordeaux UMR 5251

3. ImmunoConcEpT UMR 5164

**A**ntinuclear Antibody Screening test using HEp-2 cells is the gold standard assay for patients with a suspected auto-immune disease (American College of Rheumatology recommendation). This assay requires expertise and is time consuming for laboratory personnel. The ICAP classification provides patterns' nomenclature for globally harmonized practice, but this is not sufficient. We hypothesized that a computer-assisted interpreting tool may improve the management of auto-immune diseases' diagnostic and follow-up.

## Background

Indirect immunofluorescence assay (IFA) using human epithelial cell tumor (HEp-2) cells is the most established method for ANA screening (ANA IFA) [1] and considered the "gold standard" by the ACR [2].

The main advantages of the ANA IFA are the detection of a wide range of autoantibodies, high sensitivity and the ability to determine staining patterns and titers simultaneously. Nevertheless, the ANA IFA has several drawbacks including the use of manual microscopes for interpretation of IFA patterns, leading to a high degree of subjectivity with potential significant inter- and intra-laboratory variations [3]. With this in mind, the International Consensus on ANA Patterns (ICAP) has drawn up recommendations to achieve consensus on the nomenclature and definition of IFA pattern on HEp-2 cells. These recommendations consist in a standardized nomenclature distinguishing different fluorescence aspects from anti-cellular (AC) -0 to AC-29, as well as their interpretation. To overcome these difficulties and given the limitations of several alternative techniques (e.g. solid-phase assays), the biomedical industry has developed technological solutions in the past few years. These solutions significantly improve the automation of IFA on HEp-2 cells both with substrate and slide preparation, and microscope reading. The NOVA®View system (Inova Diagnostics, Werfen, San Diego, California) enables HEp-2 slides to be read and digital patterns to be acquired. The system uses barcode readers for full traceability, and a motorized fluorescence microscope with autofocus. A software using mathematical algorithms based on pixel intensity and light distribution at 1/80 dilution is embarked. It allows positive and negative discrimination, identification of aspects (homogeneous, speckled, centromere, nucleolar and nuclear dots), cytoplasm detection and titer determination [4].

The Immunology Laboratory of Bordeaux University Hospital is equipped with the NOVA®View system since 2015 which has greatly improved the laboratory productivity for IFA on HEp-2 cells (saving technical time, harmonizing results reporting, improving traceability, archiving...). While discrimination between positive and negative samples is satisfactory [5], the system's ability to propose fluorescence patterns is incomplete, if not inaccurate. This pitfall requires extensive proofreading and reviewing with modification by technical staff.

## Contribution

Images from previous diagnostics at Bordeaux University Hospital were collected in order to train different neural networks in terms of architecture using PyTorch (DenseNet, EfficientNet B4, ResNext50, ResNext34, ResNext18). These structures were pre-trained on ImageNet (3 color channels) and transformed to perform on only 1 color channel. These networks were combined with a novel multi-head structure to obtain a multi-label classifier.

Two separate tasks were performed : the first focusing on the nucleus of the cells and the latter on their cytoplasm. In order to do so, we used the DAPI (fluorescent stain that binds strongly to adenine-thymine-rich regions in DNA) to filter nucleus zones of the images. One main issue was the unbalanced distribution of each class : we had to do samplings with a focus on minority classes, with the use a cost function [6] that takes this imbalance into account. With the validation cohort (80-20% splits), the following performance metrics were obtained :

— Accuracy : 0.9771



- AUROC (mean over classifier heads) : 0.9751
- Hamming Loss : 0.02112
- Exact match : 0.8475
- Ranking loss : 0.004775
- Ranking Average Precision : 0.9872

Similar results were obtained with cross- validation. Same work has been done with cytoplasmic patterns.

## Outlooks

We started to work on a Generative Adversarial Network (GAN) in order to generate characteristics that will be used for non-supervised learning. We will also have a look on an autoencoder. Finally, a radiomics approach will be used to test cell by cell quantitation.

## Conclusion

We report on the methodology and establishment of a useful computer-assisted interpreting tool for indirect immunofluorescence on HEp-2 cells, representative of auto-immune diseases encountered in a Western European country. This tool paves the way for other computer-assisted interpreting tools for the practicing medical biologist or pathologist.

## Bibliographie

- [1] Pisetsky DS. Antinuclear antibody testing — misunderstood or misbegotten? *Nat Rev Rheumatol*. 2017 Aug;13(8) :495-502.
- [2] Meroni PL, Schur PH. ANA screening : an old test with new recommendations. *Ann Rheum Dis*. 2010 Aug 1;69(8) :1420-2.
- [3] Choi HW, Kwon YJ, Park JH, Lee SY, Chun S, Won EJ, et al. Evaluation of a Fully Automated Antinuclear Antibody Indirect Immunofluorescence Assay in Routine Use. *Front Immunol*. 2020 Dec 4;11 :607541.
- [4] Bizzaro N, Antico A, Platzgummer S, Tonutti E, Bassetti D, Pesente F, et al. Automated antinuclear immunofluorescence antibody screening : A comparative study of six computer-aided diagnostic systems. *Autoimmun Rev*. 2014 Mar;13(3) :292-8.
- [5] Zheng B, Li E, Zhu H, Lu J, Shi X, Zhang J, et al. Automated antinuclear immunofluorescence antibody analysis is a reliable approach in routine clinical laboratories. *Clin Chem Lab Med CCLM [Internet]*. 2017 Jan 26 [cited 2021 Oct 14];55(12).
- [6] Baruch, Emanuel Ben et al. "Asymmetric Loss For Multi-Label Classification." 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2020) : 82-91.

# AMÉLIORATION DU SUIVI CLINIQUE DE TUMEURS INTRACRÂNIENNES À L'AIDE D'ÉQUATIONS DIFFÉRENTIELLES ET DE MÉTHODES D'APPRENTISSAGE

Virginie MONTALIBET

Direction : Olivier SAUT<sup>1</sup>

Co-Direction : Annabelle COLLIN<sup>1</sup>

Encadrement : Julien ENGELHARDT<sup>2</sup>

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

2. Service de Neurochirurgie B, CHU de Bordeaux, Place Amélie Raba-Léon, 33076 Bordeaux Cédex, France

Les méningiomes représentent des tumeurs généralement non cancéreuses qui se développent dans le système nerveux central. Il est essentiel d'assurer un suivi médical régulier des patients afin d'évaluer les risques potentiels associés à la croissance de ces tumeurs, en prenant en compte des critères tels que leur évolution, leur volume et leur localisation. Actuellement, cette évaluation repose sur l'interprétation des IRM réalisées lors des examens cliniques. Toutefois, ces examens présentent des risques pour la santé du patient. Nous essaierons dans cette thèse de proposer des outils pour assister les chirurgiens dans le suivi clinique de ces patients.

## Introduction

Naissant des cellules arachnoïdiennes, les méningiomes font aujourd'hui partie des tumeurs intracrâniennes les plus fréquentes du système nerveux central [2]. Majoritairement bénignes et à croissance faible, une grande proportion de patients demeure asymptomatique tout au long sa vie. Ainsi, plus de la moitié des méningiomes sont découverts durant l'autopsie du patient [2].

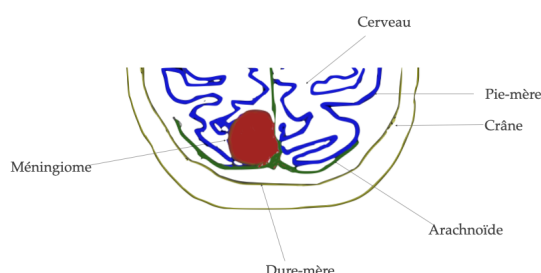


FIGURE 50.1 – Anatomie du cerveau et visualisation d'un méningiome.

Caractérisés comme des tumeurs à croissance lente, leur développement reste néanmoins variable et peut s'avérer très différent d'une tumeur à l'autre [3]. Par conséquent, ils ne présentent pas toujours un risque majeur pour la santé. L'opération chirurgicale, consistant à enlever la tumeur, pourrait être plus dangereuse et douloureuse que la présence du méningiome chez le patient. Il en est de même pour les traitements de radiothérapie, parfois lourds à suivre.

Ainsi, les patients atteints de méningiomes sont régulièrement suivis par les cliniciens jusqu'à ce que la crois-

sance de ces tumeurs présente un risque majeur pour leur santé. Afin d'évaluer la fréquence du suivi médical, les chirurgiens se basent sur de nombreuses caractéristiques incluant la croissance, le volume, la symptomatologie ou encore la position de la tumeur dans le cerveau [3].

La majorité de ces caractéristiques peut être calculée ou mesurée grâce aux imageries faites pendant l'examen du patient.

Afin d'aider les chirurgiens dans ce suivi, nous analyserons dans cette thèse trois des caractéristiques utilisées par les cliniciens pour prendre une décision : l'identification de la tumeur et sa segmentation, la croissance de celle-ci dans le temps et enfin la prédiction de son volume.

Pour cela, une base de données composée de 333 méningiomes sera utilisée. Les patients associés sont tous suivis au CHU de Pellegrin à Bordeaux.

## Segmentation

Le segmentation des tumeurs sur les images IRM a un rôle majeur. En effet, elle permet non seulement d'identifier la tumeur mais aussi l'extraction de caractéristiques importantes. À l'heure actuelle, elle est faite manuellement par les chirurgiens ou en utilisant des outils de segmentation semi-automatiques. Ces outils se basent sur des changements de contraste et/ou d'intensité pour délimiter et extraire la zone tumorale. Bien que nécessitant toujours l'intervention humaine, ces algorithmes diminuent significativement le temps d'exécution de cette tâche.

Notre objectif réside dans la mise en place d'algorithmes

totalement automatiques de segmentation. Pour cela, plusieurs architectures de réseaux de neurones ont été testées sur nos données. Des exemples de segmentations obtenues sont visibles dans la figure 50.2.

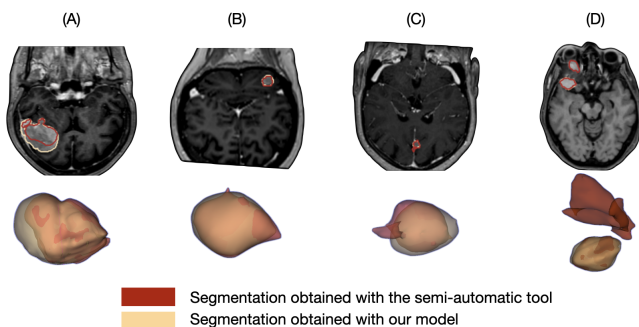


FIGURE 50.2 – Exemple de segmentations automatiques obtenues

## Croissance

Une fois la segmentation faite sur chaque examen, les caractéristiques importantes des tumeurs peuvent être extraites. Parmi elles, le volume de la tumeur. Ainsi, il est possible de visualiser l'évolution du volume de la tumeur dans le temps. Nous souhaitons alors décrire cette croissance tumorale à l'aide de modèles mathématiques. Ces modèles prennent la forme d'équations différentielles ordinaires (EDO) et seront utilisés pour "fitter" nos données. Nous présentons ici le modèle de Gompertz.

$$\begin{cases} V'(t) = V(t)\alpha e^{-\beta t} \\ V(t=0) = V_0. \end{cases}$$

Il a pour solution analytique :

$$V(t) = V_0 e^{\frac{\alpha}{\beta}(1-e^{-\beta t})}$$

avec  $V(t)$  le volume de la tumeur au temps  $t$ .

Ces modèles dépendent de paramètres qui ont été estimés à l'aide de différentes stratégies.

Les stratégies individuelles ont montré une forte dépendance aux bruits de mesures. Ces limitations nous ont donc poussés à tester une autre approche, cette fois-ci en population. Par sa forte robustesse aux bruits, cette stratégie a été utilisée pour obtenir les paramètres de chaque individu et ainsi analyser la croissance tumorale [4].

La comparaison de différents modèles nous a permis d'identifier le modèle de Gompertz comme étant celui qui s'adapte le mieux à nos données. La figure 50.3 montre les résultats obtenus par différents modèles.

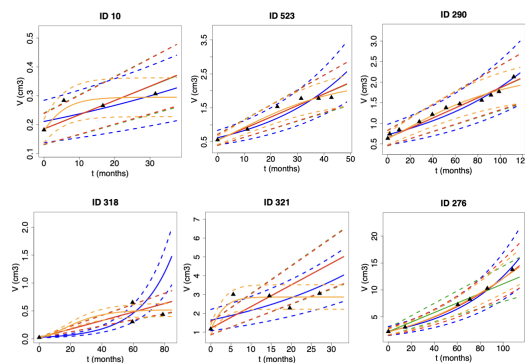


FIGURE 50.3 – Exemples de courbes obtenues. Bleu : exponentiel, Rouge : Puissance, Orange : Gompertz, Vert : Linéaire

Le modèle de Gompertz permet une meilleure compréhension de la croissance des méningiomes caractérisée par une croissance lente, suivie d'une accélération exponentielle avant de finir par se stabiliser et atteindre sa taille maximale. La prédiction d'un futur volume ou de ces différentes étapes de la croissance du méningiome serait donc particulièrement utile aux chirurgiens.

## Prédiction

Afin d'identifier les différentes phases du modèle de Gompertz sur nos données, une méthode de clustering hiérarchique a été mise en place. Ainsi, des clusters obtenus lors de la première et de la dernière visite du patient au CHU ont pu être identifiés. Le changement de cluster, caractérisé par le ralentissement de la croissance peut donc être mesuré et étudié.

Enfin, différents modèles ont été testés pour permettre la prédiction d'un volume futur.

## Bibliographie

- [1] Balafar, M. A. and Ramli, A. R. and Saripan, M. I. and Mashohor, S, Review of Brain MRI Image Segmentation Methods, 2010, Artificial Intelligence Review.
- [2] Baldi, I. and Engelhardt, J. and Bonnet, C. and Bauchet, L. and Bertheaud, E. and Grüber, A. and Loiseau, H., Epidemiology of Meningiomas, 2018, Neurochirurgie.
- [3] Pamir, M. Necmettin and Özdoğan, Koray, Chapter 15 - Current Decision-Making in Meningiomas, Handbook of Clinical Neurology, McDermott, Michael W., 2020.
- [4] Engelhardt J and Montalibet V and Saut O and Loiseau H and Collin A, "Evaluation of four tumour growth models to describe the natural history of meningiomas.", EBioMedicine., "2023 Aug;".

# OPTIMAL STOCHASTIC CONTROL AND APPLICATION ON PASSIVE TRAJECTOGRAPHY

Romain NAMYST

Direction : François DUFOUR

Co-Direction : Alexandre GENADOT

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

We consider the classical problem of localization of a target from an observer from bearing measurements. We reformulate this problem within the framework of the theory of partially observed Markov decision processes and propose a method for numerically solving this problem. Theoretical convergence of this numerical solution scheme is obtained and numerical investigations are also carried out, enabling us to recover optimal curves already suggested in the literature via other techniques.

## Problem formulation

In this paper, we consider the classical problem of localization of a target from an observer from bearing measurements, see for example [1] and the references therein. In this problem, an observer at position  $\mathbf{x}^O$  wants to infer the position  $\mathbf{x}^T$  of a static target, both in the same horizontal plane, from noisy bearing measurements, that is, from the successive noisy measurements of the angle between the observer-target direction and the north direction. This situation is illustrated in Figure 51.1.

In this setting of localization, the position of the target is observable in a statistical sense, meaning that it is possible to infer the position of the target from the measurements, even when the observer does not perform any manoeuvre, see [2]. Nevertheless, one or many manoeuvres might for instance speed up the inference process or improve its accuracy, or both. It is therefore of great interest to try to optimize the choice of this possible manoeuvres along time for the observer in order to infer the position of the target with the best possible precision.

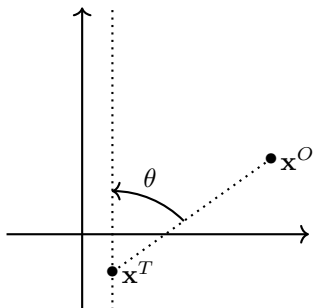


FIGURE 51.1 – An observer at point  $\mathbf{x}^O$  and the target at point  $\mathbf{x}^T$ . The bearing  $\theta$  is also displayed on the figure.

A criteria of optimisation is thus to be defined. In the literature, functions of the bearing rate have been considered since this quantity naturally arises as a lower bound for the determinant of the Fisher information ma-

trix associated to this problem, see in particular [4, 5] on this subject. This criteria has a simple interpretation : in between two successive measurements, we want to have the greatest possible difference between the two measured angles. Note that this criteria being related to the Fisher information matrix, it allows also to approach the best estimate in the sense of Fisher information : it can be viewed as an approximation of the associated Cramer-Rao lower bound.

A lot of methods have been proposed to tackle this optimization problem, often in the context of tracking, that is, when the target is itself moving. In this PhD, the author mention the fact that the theory of partially observed Markov decision processes could be use but gives raise to non tractable, or at least non efficient, numerical algorithms. Nevertheless, we are going to continue along this path in the present document, in order to face up to these practical difficulties and propose a method for partially overcoming them.

A Markov decision process, a MDP in what follows, is, roughly speaking, a discrete time Markov chain whose transitions depend on some control parameter and which are endowed with some one step reward function, depending on current state and action, or more generally on the whole trajectories of the state and action processes. The aim of the controller is then to choose the successive controls so that it maximize some cumulative reward criteria.

In practice, it happens that the transitions of the MDP depend on some unknown parameters, as in the context of adaptive control for MDP, see [6, 8], or that the true state of the MDP is only observed through some measurement function. In this latter case, we are in the context of partially observed Markov decision processes, POMDP in the sequel. In order to solve a POMDP, that is, to find what is the best cumulative reward we can expect and what are the actions to do in order to earn it, a common trick consists in enlarging the state space to some probability space often called the belief space,

see for example [3]. The new process with the belief space as the state space is a fully observed MDP that can be solved, in theory, using the classical techniques, such as the resolution of the Bellman equations if the time window is finite. However, in practice, the resolution of these equations requires to browse the entire state space which therefore has to be, in most situations, discretized.

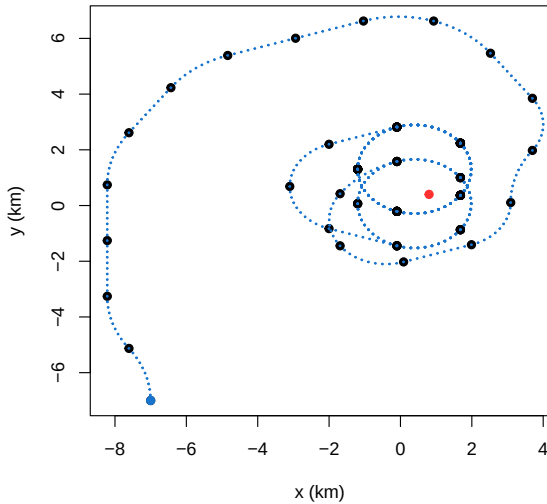


FIGURE 51.2 – a computed optimal trajectory when  $\sigma = 0.5$ . The target position is in red whereas the starting position of our ownship is in blue and then follows the blue dashed line. The black dots represent the position after each taken action. A slice of size  $H = 3$  is composed of three black dots after the starting point.

In this PhD, we try to discretized the belief space by following the dynamic of the original MDP while constructing the belief space, in order to build this one as small as possible. In our situation of bearing only tracking, this construction ends up with a space consisting in an exponential family of measures. The use of this family will allow us to propose a tractable numerical scheme to solve the Bellman equations associated to our BOT problem.

## Results

The graph in Figure 51.2 shows the optimal trajectory found using our algorithm. The quality of the method is demonstrated in Figure 51.3, we observe that the blue curve is always above the red one but that they tends to be closer as time grows. We also notice that the empirical median of the CRLB is very closed to its mean. The difference in mean between the RMSE and the CRLB after 90 time units is about 112 meters whereas it is about 46 meters for the difference between the two medians.

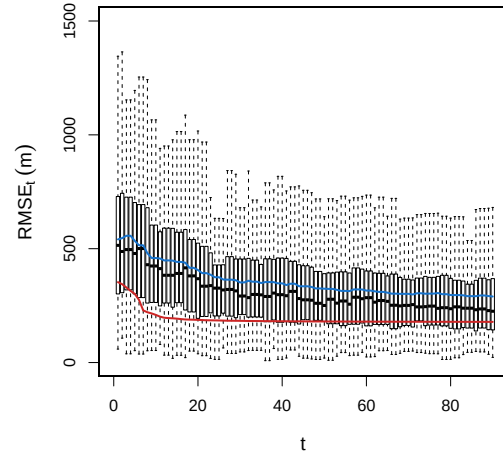


FIGURE 51.3 – boxplots of the root mean square error over time, for 100 trajectories with  $\sigma = 0.5$  with corresponding mean in blue and mean of the Cramer-Rao lower bound in red.

## Bibliographie

- [1] Zhang, H., Dufour, F., Anselmi, J., Laneuville, D. & Nègre, A. Piecewise optimal trajectories of observer for bearings-only tracking by quantization. 2017 20th International Conference On Information Fusion (Fusion). pp. 1-7 (2017)
- [2] Jauffret, C. & Pillon, D. Observability in passive target motion analysis. *IEEE Transactions On Aerospace And Electronic Systems*. **32**, 1290-1300 (1996)
- [3] Saldi, N., Yüksel, S. & Linder, T. Asymptotic optimality of finite model approximations for partially observed markov decision processes with discounted cost. *IEEE Transactions On Automatic Control*. **65**, 130-142 (2019)
- [4] Hammel, S., Liu, P., Hilliard, E., Gong, K. Optimal observer motion for localization with bearing measurements. *Computers & Mathematics With Applications*. **18**, 171-180 (1989)
- [5] Liu, P. An optimum approach in target tracking with bearing measurements. *Journal Of Optimization Theory And Applications*. **56** pp. 205-214 (1988)
- [6] Genadot, A., Dufour, F., Costa, O. Minimum Contrast Estimators for Piecewise Deterministic Markov Processes. (2023)
- [7] Costa, O., Dufour, F., Genadot, A. Adaptive Average Control for Piecewise Deterministic Markov Processes. Available At SSRN 4655338. (2023)
- [8] Hernández-Lerma, O. Adaptive Markov control processes. (Springer Science & Business Media, 2012)



# ELECTROCARDIOLOGY MODELING AFTER CATHETER ABLATIONS FOR ATRIAL FIBRILLATION

**Simone NATI POLTRI**

**Direction : Clair POIGNARD**

**Co-Direction : Annabelle COLLIN**

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

Catheter-based cardiac ablation, such as radiofrequency ablation (RFA) and pulsed electric field ablation (PFA), is the treatment of choice for atrial fibrillation (AF). However, the underlying phenomena and differences between RFA and PFA are not well understood. In my thesis, I propose a mathematical modeling of the cardiac electric signal of a cardiac domain containing an ablated area by RFA or PFA. Both types of ablation consist of the isolation of the pulmonary vein, but we describe them differently by using appropriate transmission conditions. Numerical simulations in a context of AF show that PFA and RFA lead to isolation of the pulmonary vein.

## Introduction

Cardiac arrhythmias are irregularities in the heartbeat that result in chaotic electrical waves. While most of cardiac arrhythmias are benign, some of them can directly affect the pumping function of the heart, leading then to stroke or heart failure. Isolation of the pulmonary veins by catheter ablation has become the treatment of choice for atrial fibrillation (AF). The goal is to isolate the pulmonary veins from which the fibrillation is supposed to originate by physical procedures such as thermal ablation (radiofrequency ablation (RFA) [1]), and more recently pulsed electric field ablation (PFA), which is based on nonthermal irreversible electroporation.

Despite the great interest that RFA and PFA have generated in the treatment of AF, there is still a lack of understanding – and thus modelling – of the underlying biophysical phenomena of these different therapies. On one hand, it is well known that RFA ablation leads to coagulation necrosis with complete loss of cellular and vascular architecture [2] by leaving a scar composed of a fibrotic tissue. On the other hand, PFA is known to destroy mainly the cardiomyocytes, but the tissue scaffold is preserved. Therefore, the physical properties of the cardiac tissue after RFA and PFA are very different, although they have the same goal, which is to isolate the pulmonary vein.

Well-designed mathematical modeling could help to better understand the effects of PFA on cardiac electrical wave and to develop numerical criteria for treatment evaluation. More precisely, we propose a mathematical modeling of the cardiac electrical signal of a cardiac domain containing an area ablated by RFA or PFA. We study the effects of this treated area on the propagation of the electrical wave, known to correspond to the so-called bidomain model [3] in cardiac domain, through well-adapted transmission conditions across the treated region.

## The mathematical modeling

### PFA modeling

It has been experimentally observed that PFA preserves the tissue scaffold and targets the myocardium through the nonthermal, irreversible electroporation process. After PFA, the mechanical properties (stiffness, elasticity...) of the cardiac tissue are preserved, whereas the electrical functionalities of the ablation area are altered. Most likely, PFA leads to local death of cardiomyocytes, which are then replaced by nonexcitable fibroblasts.

From a mathematical point of view, we consider an interface  $\Gamma$  to be the limit of a thin area of thickness  $\varepsilon$  as far as  $\varepsilon$  tends to zero, and we study the formal transmission conditions of the electric potentials, derived from a formal asymptotic development.

Therefore, we propose to model the electrical effect of PFA by a continuity of both the extracellular potential and the extracellular normal flux, while assuming that the intracellular potential is isolated, thus using a homogeneous Neumann boundary condition. Roughly speaking, these conditions arise from the zero order of the asymptotic analysis, in which the small parameter  $\varepsilon$  is the thickness of the electroporated (EP) area and the intracellular conductivity tensor is assumed to scale proportionally to a parameter  $\varepsilon^2$ . In order to simplify the analysis, the non-linear ionic current, that plays a fundamental role in the reaction term of the bidomain model, is linearized inside the EP area.

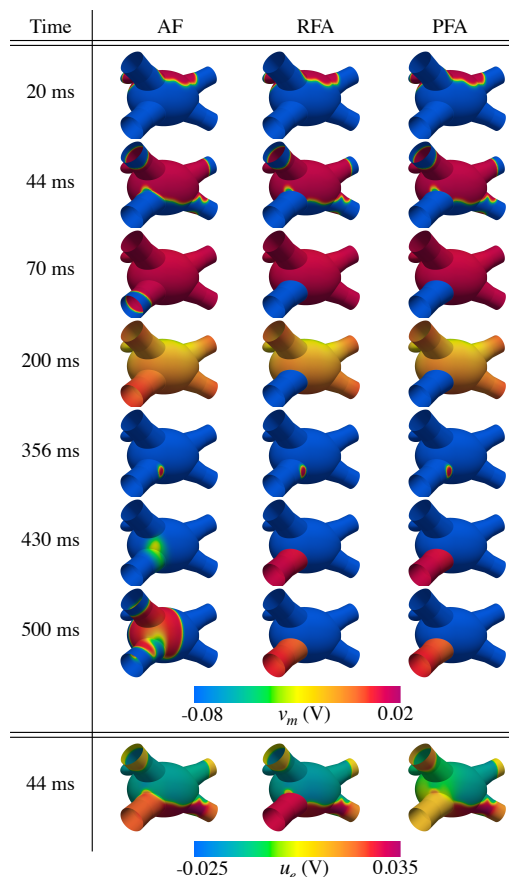
The formal asymptotic analysis and the derivation of approximated transmission conditions at any order, are completed by a proof of the expansion and of estimates for the errors of the potentials between the exact solution and the truncated solution at a generic order  $N$ . Convergence tests have been done and mathematical estimates has been numerically validated. This Section briefly resumes a first part of my thesis work.

## RFA modeling

To model the effect of RFA, we consider the well-known Kedem-Katchalsky transmission conditions – initially introduced in [4]. These conditions are defined through a coefficient  $\alpha$ , a positive constant homogeneous to a surface conductance. This parameter  $\alpha$  is crucial because it is responsible for whether or not the transmembrane potential wave can overcome the  $\Gamma$  interface. In particular, when  $\alpha = 0$ , it is equivalent to perfect isolation of the pulmonary vein. In the asymptotic regime  $\alpha \gg 1$ , the potentials become asymptotically continuous. The Kedem-Katchalsky can then be seen as a penalty term which weakly enforces the continuity of the potential through the interface  $\Gamma$ .

## Numerical Illustrations

To compare the effects of the two ablations considered, a simulation corresponding to atrial fibrillation before ablation is proposed. Physiologically, the depolarization wave that triggers the heartbeat is initiated in the sinoatrial node in the right atrium. It then propagates to the left atrium via two electric pathways : the fastest leading to Bachmann's bundles (BB) and the second to the fossa ovalis (FO). The protocol and parameters were set in [5].



**FIGURE 52.1** – First column : before ablation. Second column : successful RFA with  $\alpha = 10^{-4}$ . Third column : successful PFA. Lines 1 to 7 : Snapshots of transmembrane potential  $v_m$ . Last line : Snapshot of extracellular potential  $u_e$ .

The first column of Figure 52.1 shows time snapshots of

the transmembrane potential before ablation. One first sees the healthy depolarization occurring during the first 90ms – see 20, 44, and 70ms snapshots – followed by the healthy repolarization – see 200ms snapshot – and by the second stimulus illustrating a pathological area of one of the pulmonary veins – see 356ms snapshot – that triggers a pathological wave that unfolds in the left atrium – see 430 and 500ms snapshots.

The second column of Figure 52.1 gives the transmembrane potential corresponding to successful RFA, namely K-K conditions with  $\alpha = 10^{-4}$ . Because the two regions are nearly decoupled, the pulmonary vein is well isolated : there is no entry of the wave into the left atrium, see 356, 430, and 500ms snapshots. Uncoupling is also seen in the evolution of the extracellular potential, shown in the last line of the same figure at time 44ms. The third column of Figure 52.1 corresponds to a successful PFA. One can first see that there is no entry of the transmembrane wave, showing the success of the ablation. However and contrary to RFA (second column), the PFA maintains the continuity of the extracellular potential  $u_e$ , see the last line of Figure 52.1.

## Bibliographie

- [1] Joseph, Wout, et al. "Determination of the duty cycle of WLAN for realistic radio frequency electromagnetic field exposure assessment." *Progress in Biophysics and Molecular Biology* 111.1 (2013) : 30-36.
- [2] Bulvik, Baruch E., et al. "Irreversible electroporation versus radiofrequency ablation : a comparison of local and systemic effects in a small-animal model." *Radiology* 280.2 (2016) : 413-424.
- [3] Tung, Leslie. A bi-domain model for describing ischemic myocardial dc potentials. Diss. Massachusetts Institute of Technology, 1978.
- [4] Kedem, O., and A. Katchalsky. "A physical interpretation of the phenomenological coefficients of membrane permeability." *The Journal of general physiology* 45.1 (1961) : 143-179.
- [5] Schenone, Elisa, Annabelle Collin, and Jean-Frédéric Gerbeau. "Numerical simulation of electrocardiograms for full cardiac cycles in healthy and pathological conditions." *International journal for numerical methods in biomedical engineering* 32.5 (2016) : e02744.

# SIMULATIONS NUMÉRIQUES STANDARDISÉES DE DISPOSITIFS DE STIMULATION ÉLECTRIQUE CARDIAQUE

Valentin PANNETIER

Direction : Yves COUDIERE  
Encadrement : Michael LEGUÈBE

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

**L**es maladies cardio-vasculaires sont la première cause de mortalité dans le monde. Dans certains cas d'anomalie de l'activité électrique cardiaque, l'implantation d'un dispositif de stimulation électrique est nécessaire. La prise en compte de ces dispositifs est assez peu développée dans les modèles mathématiques utilisés en électrophysiologie cardiaque actuellement. L'objectif de cette thèse s'inscrit dans celui plus large du projet européen SimCardioTest<sup>12</sup>. Novatrice, elle a pour but d'explorer de nouveaux horizons et d'améliorer la compréhension de deux phénomènes biologiques observés cliniquement et expérimentalement : l'effet de seuil reliant la stimulation et le tissu excitable, et l'impédance de contact à la surface des électrodes de stimulation. La validation des différents modèles 0D et 3D suit l'approche standardisée V&V40 à travers une confrontation aux données expérimentales.

## Introduction

**Contexte.** Le projet européen SimCardioTest vise à démontrer l'intérêt et à encourager l'utilisation de la simulation numérique pour le développement de dispositifs médicaux. Conduit selon le standard V&V40, ce projet permettra à terme d'accélérer de façon significative certaines étapes de conception ou d'étude, en remplaçant des tests *in vivo* par des tests *in silico*, et en assurant la crédibilité des modèles et des logiciels utilisés. Les avancées de cette thèse sont le fruit du travail effectué conjointement par l'équipe Inria CARMEN, le fabricant de dispositifs de stimulation cardiaque MicroPort CRM, et l'institut de rythmologie cardiaque IHU Liryc avec lesquels les expérimentations animales ont été conduites.

**Électrophysiologie cardiaque.** Le tissu cardiaque est riche de phénomènes biologiques multi-échelles. Des anomalies de ce tissu mènent bien souvent à des problèmes médicaux impactant le cycle de contraction cardiaque. Lorsque le cœur n'assure plus ce cycle, il peut être nécessaire d'implanter un dispositif de stimulation extérieur. La prise en compte de ce dernier n'est traditionnellement pas étudiée dans les modèles mathématiques, où les stimulations sont résumées à de simples conditions aux bords. Que ce soit sous la forme d'injection de courant ou d'application de tension, la forme du signal, l'amplitude ou la durée de la stimulation ne sont pas intégrées. La confrontation à des données réelles de mesure est dès lors compliquée.

**Stimulations.** Deux phénomènes biologiques sont au centre de cette thèse, l'effet de seuil et l'impédance de

contact. Lorsqu'une stimulation est appliquée, elle ne mène pas toujours à la propagation d'un potentiel d'action. Lorsque c'est le cas, une contraction cardiaque est induite par la stimulation, et on parle de « capture ». Cette propriété est intrinsèquement liée à l'amplitude et à la durée de la stimulation, mais également à l'état du tissu et à la géométrie de la sonde utilisée. Dans le plan de stimulation durée-amplitude, la courbe qui délimite la région de capture a été étudiée par Louis Lapicque au début du XX<sup>e</sup> siècle [1]. Une étude approfondie de cette courbe de seuil nécessite alors une connaissance fine de la stimulation appliquée par le dispositif, mais aussi la compréhension du phénomène d'impédance de contact. En effet, lorsqu'une électrode est plongée dans un milieu riche en espèces ioniques, une couche ionique de Debye se forme à la surface des électrodes. Selon les espèces présentes cette couche a un effet résistif et capacitif, et elle impacte l'énergie reçue par le tissu. Cet effet est d'autant moins négligeable dans le cas d'une sonde de stimulation, puisqu'une des électrodes est directement implantée dans le tissu cardiaque.

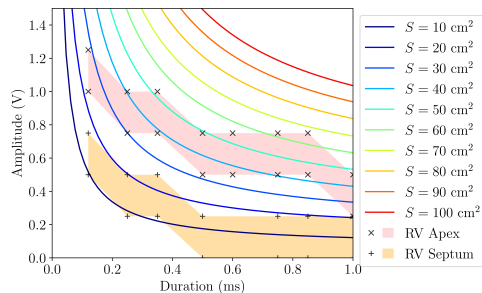
## Modèles

**Modèle de stimulateur.** Deux catégories de modèles ont été étudiées durant cette thèse : 0D *i.e.* sans dimension spatiale, et 3D. Cependant, ils sont tous basés sur un même système d'équations différentielles ordinaires. Il traduit mathématiquement le circuit interne du dispositif de stimulation cardiaque étudié. Un cycle de stimulation se décompose en plusieurs sous-parties distinctes discontinues en temps, dont les deux principales sont *le pulse* et *la décharge*, ce qui rend l'analyse plus ardue.

<sup>12</sup>. [www.simcardiotest.eu](http://www.simcardiotest.eu)

**Modèles équivalents de contact.** Une étude préliminaire a été menée pour identifier les paramètres des circuits électriques équivalents reproduisant la couche de Debye à la surface des électrodes. Certains modèles issus de la littérature sont formulés en série ou font intervenir des composants non idéaux et fréquentiels, ce qui rend l'identifiabilité numérique des paramètres difficile ou impossible. En confrontant ces modèles aux données expérimentales, nous avons retenu le circuit équivalent formé d'une résistance en parallèle avec un condensateur – meilleur en terme d'identifiabilité et de propriétés qualitatives.

**Modèles 0D.** Une première version du modèle a été présentée à la conférence FIMH 2023 [2]. Le modèle 0D proposé a été comparé à différentes données expérimentales, notamment pour la phase de calibration des paramètres, et une étude succincte de sensibilité locale a été effectuée. La figure 53.1 illustre l'influence d'un des paramètres du modèle sur la courbe de Lapique, ainsi que les zones de seuil expérimentaux *ex vivo* sur cœur de brebis.

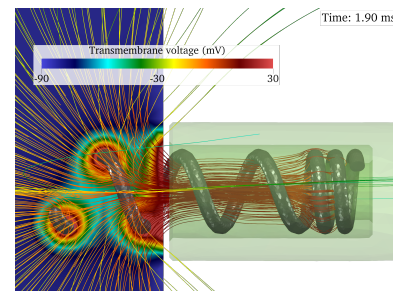


**FIGURE 53.1** – Influence sur la courbe de Lapique du paramètre de mise à l'échelle  $S$  dans le modèle 0D à une cellule. Les zones colorées sont les zones expérimentales de seuil durant la première expérimentation animale. Source [2].

Ce modèle a ensuite été enrichi et fera l'objet d'une publication ultérieure (CinC 2024) sous la forme d'un modèle 0D à deux cellules. Ces modèles sont beaucoup moins coûteux numériquement qu'un modèle 3D complet. Ils constituent un outil important dans la compréhension des différents phénomènes et surtout dans l'exploration de nouvelles pistes de recherche.

**Modèle 3D.** En parallèle, un modèle 3D a aussi été développé sur la base du modèle de propagation bidomaine utilisé en électrophysiologie cardiaque. Nous y avons rajouté un milieu de propagation extracardiaque, et nous avons couplé le tout au circuit électrique du stimulateur à travers des conditions de bords non usuelles. Inspiré des travaux [3], ce modèle fera l'objet d'une présentation au CANUM 2024 et d'un article théorique sur le caractère bien posé [4]. Ce modèle, plus complexe et plus coûteux numériquement que les modèles 0D, est utilisé sur la plateforme propulsée par InSilicoTrials dans le cadre de SimCardioTest. Cependant, il nécessite d'être calibré à partir de données expérimentales volumineuses et multiples. Ces travaux dépassent le cadre de la thèse et constituent un projet de recherche à part entière. La figure 53.2 est un exemple de résultat du modèle 3D. Le maillage de sonde, et en particulier la vis

déployable avec fixation active, est basé sur une géométrie de sonde MicroPort. Les paramètres de conductivité sont issus de la littérature, et le système d'équations est résolu à l'aide du code de calcul Ceps par la méthode des Éléments Finis.



**FIGURE 53.2** – Exemple de simulation 3D. Les lignes de champ électrique sont impactées par les différences de conductivité entre le tissu et le sang, et la tension transmembranaire à la fin du pulse est illustrée en couleur.

## Conclusion

Plusieurs modèles ont été développés au cours de cette thèse. Ils permettent l'étude approfondie de l'effet de seuil lors d'une stimulation, et mettent en lumière l'importance de l'impédance de contact et de la forme de la stimulation appliquée. Ces différents modèles ont été plusieurs fois confrontés à des données expérimentales et de nombreux paramètres ont été calés à l'aide de ces mêmes données. La démarche constante de vérification et de validation permet de crédibiliser les simulations numériques dans le développement de dispositifs médicaux, en particulier pour les dispositifs de stimulation cardiaque.

## Bibliographie

- [1] Lapique, L., 1909. **Définition expérimentale de l'excitabilité.** Soc. Biol.
- [2] Pannetier, V., Leguèbe, M., Coudière, Y., Walton, R., Dhiver, P., Feuerstein, D. and Amaro, D., 2023, June. **Modeling Cardiac Stimulation by a Pacemaker, with Accurate Tissue-Electrode Interface.** FIMH 2023.
- [3] Estévez Schwarz, D. and Tischendorf, C., 2000. **Structural analysis of electric circuits and consequences for MNA.** Int. J. Circ. Theor. App.
- [4] *En préparation* : Coudière, Y., Leguèbe, M., Pannetier, V. **Bidomain model coupled to a realistic multi-electrode device in cardiac electrophysiology.**

Les auteurs ont reçu des fonds du programme de R&I Horizon 2020 de l'UE dans le cadre des accords de subvention numéro 101016496 [SimCardioTest].



# LOIS DE PAROI À APPRENTISSAGE PROFOND POUR SIMULATIONS AÉRODYNAMIQUES

**Michele ROMANELLI**

**Direction : Héroïse BEAUGENDRE**

**Encadrement : Michel BERGMANN, Denis SIPP, Samir BENEDDINE, Ivan MARY**

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

*Aerodynamics, Aeroelasticity and Acoustics Department (DAAA), ONERA, Polytechnic Institute of Paris, F-92190 Meudon, France*

Cette thèse vise à proposer une loi de paroi basée sur l'apprentissage profond pour la modélisation des écoulements pariétaux dans les simulations RANS. Différentes approches, basées sur des réseaux de neurones, ont été proposées au cours de l'étude afin d'améliorer la précision et réduire les coûts de calcul du modèle proposé. Les résultats démontrent une précision élevée, même dans des configurations non rencontrées lors de l'entraînement des réseaux neuronaux composant le modèle. Des perspectives d'amélioration de la capacité prédictive du modèle, notamment pour prédire la séparation de l'écoulement, sont discutées.

## Introduction

La disponibilité de lois de paroi fiables et précises est l'un des principaux défis de la CFD moderne. Les modèles de paroi sont obligatoires pour l'application des simulations aux grandes échelles de turbulence (LES) pour des écoulements représentatifs d'applications aéronautiques [1]. Ils sont également parfois indispensables pour les simulations Reynolds-Averaged-Navier-Stokes (RANS) soit lors de l'utilisation de méthodes de frontières immergées (IBM) [2], soit pour la mise en place de calculs légers utiles aux études paramétriques, au contrôle des écoulements ou à l'optimisation de forme, pour lesquels de nombreuses simulations sont souvent nécessaires et où le coût cumulé des calculs RANS peut devenir rédhibitoire. Dans ce contexte, les lois de paroi permettent d'alléger les calculs en remplaçant la résolution coûteuse de la couche limite (typiquement une zone maillée très finement) par un modèle.

Alors que les approches basées sur l'apprentissage profond commencent à être utilisées dans le domaine des lois de paroi, les travaux et les essais existants sont encore incomplets et une méthodologie claire n'a pas encore émergé [3]. Cependant, au regard de l'étendue des phénomènes qui peuvent être modélisés grâce à la flexibilité des réseaux de neurones, ces méthodes montrent un potentiel indéniable dans leur application pour la modélisation des écoulements pariétaux.

### Objectifs scientifiques

L'objectif de recherche est de proposer une loi de paroi basée sur des algorithmes d'apprentissage profond applicables dans le cadre de simulations Reynolds-Averaged-Navier-Stokes (RANS), surpassant les modèles classiques en terme de précision et de généralité. De plus, le caractère exploratoire de cette étude vise à identifier les grandeurs d'entrée qui permettraient une modélisation plus fiable du développement de la couche limite.

## Démarche et principaux résultats

### Approche classique

Une approche initiale consiste à entraîner un réseau neuronal en utilisant des données résolues jusqu'à la paroi pour apprendre à reconstruire des profils de vitesse adimensionnés,  $u^+$ , afin de modéliser l'évolution de la couche limite en proche paroi [4]. En se conformant au jeu de données des lois de paroi analytiques, l'évolution de la vitesse est définie en fonction de la distance à la paroi et du gradient de pression, respectivement  $y^+$  et  $p^+$ , deux grandeurs adimensionnées selon le cisaillement visqueux pariétale  $\tau_w$ , à l'aide d'une vitesse caractéristique de frottement

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}}, \quad (54.1)$$

estimée itérativement à travers un algorithme de type Newton. Cette approche se réduit donc à imposer une condition aux limites de type Dirichlet provenant de la région modélisée en proche paroi pour le calcul RANS à résoudre. Les cellules modélisées appartiennent en fait au stencil numérique utile au calcul RANS de la zone résolue.

### Approche Dirichlet-to-Neumann map

Afin de réduire le coût de calcul de l'estimation du cisaillement visqueux, une nouvelle approche a été proposée afin de remplacer le procédé itératif par une approche basée sur les données. Inspirée par le concept de *Dirichlet-to-Neumann map* [5], cette méthode impose directement le flux visqueux à travers l'interface entre la région modélisée et le calcul RANS. La dérivée normale à la paroi du champ de vitesse est donc fixée à l'interface entre la zone modélisée et l'écoulement résolu. Cela remplace la condition aux limites de Dirichlet de l'approche classique par une condition



de Neumann. L'évaluation de la dérivée du champ de vitesse est obtenue à l'aide d'un modèle composé par deux réseaux neuronaux interconnectés,  $h_{NN}$  et  $f_{NN}$ , comme montré par l'équation :

$$\frac{\partial u}{\partial y} = \frac{u_{\tau}^2}{\nu} \frac{\partial u^+}{\partial y^+} = \frac{\nu}{y^2} h_{NN}^2 f_{NN}. \quad (54.2)$$

Le premier consent une estimation du cisaillement visqueux pariétale, à travers la vitesse de frottement  $u_{\tau}$  ; le deuxième permet l'évaluation de la dérivée normale à la paroi du champ de vitesse  $\frac{\partial u^+}{\partial y^+}$ .

### Données d'entraînement et évaluation

Les deux réseaux neuronaux sont entraînés à l'aide de données de simulations RANS entièrement résolues d'écoulements turbulents le long de différentes géométries de bosse bidimensionnelle. Différentes conditions d'écoulement, jusqu'aux limites du décollement, sont prises en compte en variant le nombre de Reynolds  $Re$  et la hauteur de la bosse  $h$  (ce qui modifie les gradients de pression sur la paroi). La figure 54.1 montre l'ensemble des cas sélectionnés pour l'entraînement des réseaux neuronaux et pour leur évaluation dans des conditions d'interpolation par rapport aux paramètres  $(Re, h)$ .

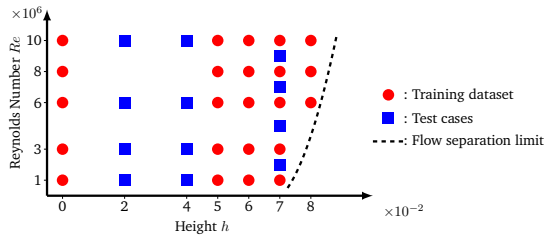


FIGURE 54.1 – Données d'entraînement et de test obtenues à partir de différentes combinaisons du nombre de Reynolds  $Re$  et de la hauteur  $h$  de la bosse.

### Résultats

Des simulations obtenues avec la modélisation pariétale sont enfin comparées à des calculs résolus jusqu'à la paroi pour les différentes configurations abordées. Un exemple des performances du modèle proposé est fourni en figure 54.2 pour un écoulement sur une bosse. Il a été obtenu pour une combinaison de paramètres d'écoulement (nombre de Reynolds) et de géométrie (hauteur de bosse) qui n'a jamais été rencontrée pendant l'entraînement. Dans l'ensemble, ce modèle basé sur les données démontre une excellente robustesse et une très bonne précision. Il permet de reproduire le coefficient de frottement  $C_f$  de référence avec une erreur de seulement quelques pourcents pour tous les cas test considérés.

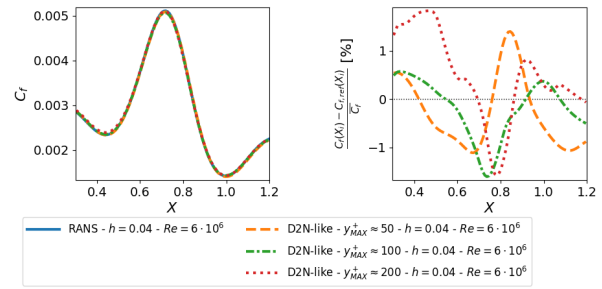


FIGURE 54.2 – Cas test de l'écoulement sur une bosse ( $h = 0.04$  et  $Re = 6 \cdot 10^6$ ). Coefficient de frottement  $C_f$  et son erreur par rapport à la simulation RANS résolue à la paroi. L'erreur est normalisée avec la valeur moyenne du  $C_f$  de référence. Hauteur de la région modélisée :  $y_{MAX}^+ \approx 50, 100, 200$ .

Afin de tester la robustesse et la généralité du modèle sur une géométrie très différente des cas d'entraînement, la loi de paroi obtenue a été également testée sur un profil NACA 4412. La précision du modèle est consistante à celle montrée sur la bosse bidimensionnelle.

### Perspectives

Malgré les bonnes performances obtenues par le modèle au sein des calculs RANS abordés, des traitements plus avancés pourraient être proposés pour améliorer les capacités prédictives des écoulements pariétaux qui présentent de forts gradients de pression et étendre le domaine de validité du modèle à des problèmes plus complexes, en se rapprochant progressivement de la prédiction de la séparation de l'écoulement.

### Bibliographie

- [1] Piomelli, U. (2008). Wall-layer models for large-eddy simulations. Progress in Aerospace Sciences, 44(6), 437-446.
- [2] Mittal, R. and Iaccarino, G. (2005). Immersed boundary methods. Annual Review of Fluid Mechanics, 37, 239-261.
- [3] Zhou, Z. and He, G. and Yang, X. (2021). Wall model based on neural networks for les of turbulent flows over periodic hills. Physical Review Fluids, 6(5), 1-30.
- [4] Romanelli, M. and Beneddine, S. and Mary I. and Beaugendre, H. and Bergmann M. and Sipp, D. (2023). Data-driven wall models for Reynolds Averaged Navier-Stokes simulations. International Journal of Heat and Fluid Flow, 99, 109097.
- [5] Sylvester, J. and Uhlmann, G. (1990). The Dirichlet to Neumann map and applications. Inverse Problems in Partial Differential Equations, 42, 101.

# THE LITTLEWOOD PROBLEM AND NON HARMONIC FOURIER SERIES

Chadi SABA

Direction : Karim Kellay

Co-Direction : Philippe Jaming

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

We give some inequalities about  $L^1$ -norms of non-harmonic trigonometric polynomials. The first one is a quantitative version of a result by Nazarov which also covers the solution of Littlewood conjecture by Mc Gehee, Pigno and Smith (we also obtained similar inequalities for quadratic frequencies, lacunary sequences and frequencies having multidimensional structure). We end with an  $L^1$ -analogue of Kahane's result on the  $L^2$ -norm of sparse trigonometric polynomials, which we also apply on the Schrödinger equation.

## $L^2$ -Theory

The aim of this note is to give a control of coefficients of a non-harmonic (non-periodic) trigonometric polynomial  $\sum a_k e^{2i\pi\lambda_k t}$  in terms of its  $L^1$ -norm or its Besicovitch  $\mathcal{B}^1$ -norms. Recall that when  $p < +\infty$ , the Besicovitch  $\mathcal{B}_p$ -norms are defined by

$$\|\Phi\|_{\mathcal{B}_p}^p = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{[-T/2, T/2]} |\Phi(x)|^p dx.$$

Those norms are the right substitute to  $L^p([-1/2, 1/2])$ -norms to investigate non-harmonic trigonometric polynomials and their limits (that is, Besicovitch-almost periodic functions).

Let us start by recalling some facts about  $L^2$ -theory. First, note that the problem is trivial for harmonic trigonometric polynomials ( $\lambda_k = k$ ) when the interval of integration has length at least 1, since the exponentials are orthogonal in  $L^2([-1/2, 1/2])$ , so the estimate of the  $L^2$ -norm is given by Parseval's relation. The same is true for non-harmonic trigonometric polynomial in the  $\mathcal{B}^2$ -Besicovitch norm, and the result is due to Ingham [1], who gave lower and upper bounds for  $L^2([-T, T])$ -norms of (non-harmonic) trigonometric polynomials :

**Théorème 6** (Ingham [1]). *Let  $\gamma > 0$  and  $T > \frac{1}{\gamma}$ . Then there exists constants  $0 < A_2(T, \gamma) \leq B_2(T, \gamma)$  such that*  
 – for every sequence of real numbers  $\Lambda = \{\lambda_k\}_{k \in \mathbb{Z}}$  verifying  $\lambda_{k+1} - \lambda_k \geq \gamma$ ;  
 – for every sequence  $(a_k)_{k \in \mathbb{Z}} \in \ell^2(\mathbb{Z}, \mathbb{C})$ ,

$$\begin{aligned} A_2(T, \gamma) \sum_{k \in \mathbb{Z}} |a_k|^2 &\leq \frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k \in \mathbb{Z}} a_k e^{2i\pi\lambda_k t} \right|^2 dt \\ &\leq B_2(T, \gamma) \sum_{k \in \mathbb{Z}} |a_k|^2. \end{aligned}$$

Note that  $A_2(T, \gamma), B_2(T, \gamma)$  are explicit constants (see [2]). Ingham has also shown that the upper bound is

valid for any  $T > 0$  while the lower bound may not be true for  $T \leq \frac{1}{\gamma}$ . Kahane has shown that this condition can be lifted if  $\lambda_{k+1} - \lambda_k \rightarrow +\infty$  when  $k \rightarrow \pm\infty$  :

**Théorème 7** (Kahane [3]). *Let  $\Lambda = \{\lambda_k\}_{k \in \mathbb{Z}}$  such that  $\lambda_{k+1} - \lambda_k \rightarrow +\infty$  when  $k \rightarrow \pm\infty$ . Then, for every  $T > 0$ , there exists positive constants  $0 < A_2(T, \Lambda) \leq B_2(T, \Lambda)$  such that*

$$\begin{aligned} A_2(T, \Lambda) \sum_{k \in \mathbb{Z}} |a_k|^2 &\leq \frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k \in \mathbb{Z}} a_k e^{2i\pi\lambda_k t} \right|^2 dt \\ &\leq B_2(T, \Lambda) \sum_{k \in \mathbb{Z}} |a_k|^2 \end{aligned}$$

holds for every sequence  $(a_k)_{k \in \mathbb{Z}} \in \ell^2(\mathbb{Z}, \mathbb{C})$  and every  $N \geq 1$ .

The explicit constants were later obtained by Haraux (see e.g. [2]). Those inequalities have found many applications in control theory, primarily to obtain  $L^2$ -observability estimate of certain PDEs, including the Schrödinger equation.

## $L^1$ -Theory

The first task is to obtain an  $L^1$ -version of Ingham's Inequality. An  $L^1 - \ell^\infty$  estimate was obtained by Ingham [1] and a much more evolved  $L^1$  to weighted  $\ell^1$  inequality was obtained by Nazarov, inspired by the proof of Littlewood's conjecture by McGehee-Pigno-Smith [6, 7]. The result is the following :

**Théorème 8.** *Let  $(\lambda_k)_{k \in \mathbb{Z}}$  be an increasing sequence of real numbers such that there exists  $\gamma > 0$  for which  $\lambda_{k+1} - \lambda_k \geq \gamma$  for every  $k$ . Let  $(a_k)_{k \in \mathbb{Z}}$  be a sequence of complex numbers.*

In Nazarov's inequality the sum starts at 0 and may fail for symmetric sums. Also its validity for  $T = \frac{1}{\gamma}$  is an open question. Further Nazarov did not provide an

estimate of the constant  $C(\gamma, T)$ . However, his proof can be modified to establish quantitative bounds. In the case  $\gamma = 1$ , we gave a quantitative version of Nazarov's result in [4] for  $T$  large enough, and in [2] for  $T$  near 1; we state both results in the following theorem

**Théorème 9.** *Let  $\lambda_0 < \lambda_1 < \dots < \lambda_N$  be distinct real numbers and  $a_0, \dots, a_N$  be complex numbers. Then*

1. *we have*

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k=0}^N a_k e^{2i\pi\lambda_k t} \right| dt \geq \frac{1}{26} \sum_{k=0}^N \frac{|a_k|}{k+1}.$$

2. *If further  $a_0, \dots, a_N$  all have modulus larger than 1,  $|a_k| \geq 1$  then*

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k=0}^N a_k e^{2i\pi\lambda_k t} \right| dt \geq \frac{4}{\pi^3} \ln N.$$

3. *Assume further that for  $k = 0, \dots, N-1$ ,  $\lambda_{k+1} - \lambda_k \geq 1$ , then, for every  $T > 1$ , there exists a constant  $C(T)$  such that, for every  $a_0, \dots, a_N \in \mathbb{C}$ ,*

$$\frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k=0}^N a_k e^{2i\pi\lambda_k t} \right| dt \geq C(T) \sum_{k=0}^N \frac{|a_k|}{k+1}.$$

Moreover,

- (a) *for  $T \geq 72$  we can take  $C(T) = \frac{1}{122}$ ;*  
(b) *for  $1 < T \leq 2$ ,  $C(T) = O((T-1)^{15/2})$ .*

Note that this result covers multiple results (Mc Gehee, Pigno and Smith, Hudson and Leckband and Stegeman's Theorems [4]). This result is sufficient to partially extend Theorem 7 to the  $L^1$ -setting. The only thing that would be missing is that in Theorem 7, there is no minimal time needed thanks to Kahane's extension of Ingham's inequality. However, so far this is unknown in the  $L^1$ -case and our second result is precisely to prove this :

**Théorème 10.** *Let  $\Lambda = (\lambda_k)_{k \in \mathbb{Z}}$  be an increasing sequence with  $\lambda_{k+1} - \lambda_k \rightarrow +\infty$  when  $k \rightarrow \pm\infty$ . Then, for every  $T > 0$ , there exist a constant  $\tilde{A}_1(T, \Lambda) > 0$  depending on  $T$  and  $\Lambda$  only such that, if  $(a_k)_{k \in \mathbb{N}} \subset \mathbb{C}$  be a sequence of complex numbers, and  $N \geq 1$ , then*

$$\frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k=0}^N a_k e^{2i\pi\lambda_k t} \right| dt \geq \tilde{A}_1(T, \Lambda) \sum_{k=0}^{\infty} \frac{|a_k|}{1+k}.$$

*If further  $\sum_{k \in \mathbb{Z}} \frac{1}{1+|\lambda_k|}$  converges, then there also exists a constant  $A_1(T, \Lambda)$  such that, for every  $(a_k)_{k \in \mathbb{Z}} \subset \mathbb{C}$  and every  $N \geq 1$ ,*

$$\frac{1}{T} \int_{-T/2}^{T/2} \left| \sum_{k=-N}^N a_k e^{2i\pi\lambda_k t} \right| dt \geq A_1(T, \Lambda) \max_{k=-N, \dots, N} |a_k|.$$

## 1-Periodic Schrödinger equation

**Théorème 11.** *Let  $u$  be a weak solution of the Schrödinger equation*

$$\begin{cases} i\partial_t u(t, x) = \frac{1}{2\pi} \partial_x^2 u(t, x) & t \in \mathbb{R}, x \in \mathbb{T} \\ u_0 = u(0, x) & x \in \mathbb{R} \end{cases},$$

*with initial value  $u_0 \in A(\mathbb{T}) = \{f \in L^1(\mathbb{T}) : \|f\|_{A(\mathbb{T})} = \sum_{k \in \mathbb{Z}} |a_k(f)| < +\infty\}$ , where  $a_k(f)$  are the Fourier coefficients of  $f$ . Let  $t_0 \in \mathbb{R}$  and  $x_0 \in \mathbb{T}$ . Then*

1. *For  $a \in \mathbb{R} \setminus \mathbb{Z}$ , for every  $T > 0$  there exists a positive constant  $C$  depending on  $a$  and  $T$  such that*

$$\frac{1}{T} \int_0^T |u(t_0 + t, x_0 + at)| dt \geq C \sum_{k \in \mathbb{Z}} \frac{|a_k|}{1+|k|}.$$

2. *If  $a \in \mathbb{Z}$  then there exists  $u_0$  such that  $u(t_0 + t, x_0 + at) = 0$  for all  $t$ . In particular, the last inequality fails.*

## Bibliographie

- [1] A. E. INGHAM Some trigonometrical inequalities with applications to the theory of series. Math. Z. **41** (1936), 367–379.
- [2] PH. JAMING & C. SABA From Ingham to Nazarov's inequality : a survey on some trigonometric inequalities. arXiv :2311.17714 [math.CA]
- [3] J.-P. KAHANE Pseudo-périodicité et séries de Fourier lacunaires. Ann. Sci. Ecole Norm. Sup. (3) **79** (1962), 93–150.
- [4] PH. JAMING, K. KELLAY & C. SABA The Littlewood problem and non-harmonic Fourier series. Math. Ann. (2023).
- [5] F. L. NAZAROV On a proof of the Littlewood conjecture by McGehee, Pigno and Smith. St. Petersburg Math. J. **7** (1996), no. 2, 265–275.
- [6] O. C. MCGEHEE, L. PIGNO & B. SMITH Hardy's inequality and the  $L^1$  norm of exponential sums. Ann. of Math. (2) **113** (1981), no. 3, 613–618.
- [7] G. H. HARDY & J. E. LITTLEWOOD A new proof of a theorem on rearrangements. J. London Math. Soc. **23** (1948), 163–168.
- [8] PH. JAMING, K. KELLAY, C. SABA & Y. WANG On  $L^1$ -norms for non harmonic trigonometric polynomials with sparse frequencies. (to appear).

# APPROCHES GALERKIN DISCONTINUES SPATIO-TEMPORELLES SUR DES MAILLES HIÉRARCHIQUES POUR LES ÉQUATIONS DE NAVIER-STOKES INCOMPRESSIBLES

Alexis TARDIEU

Direction : Angelo IOLLO

Co-Direction : Afaf BOUHARGUANE

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

Les problèmes de mécanique des fluides se posent dans divers secteurs : énergie, aéronautique et spatial, environnement... mais aussi en biologie et santé. Pour des raisons de dimensionnement d'équipements, de prévention ou de diminution des coûts, on recourt à des simulations numériques. Elles consistent à reproduire aussi fidèlement que possible les phénomènes physiques étudiés à travers des modèles et des équations. Cependant, ces simulations sont très coûteuses en temps et en ressources de calcul. Par exemple, l'étude d'un écoulement d'air à travers une éolienne pendant quelques minutes peut nécessiter plusieurs semaines de simulation sur des super-calculateurs. Cette thèse concourt donc à un objectif simple dans son énoncé : l'écriture de schémas numériques précis, efficaces et aux coûts de calcul maîtrisés.

## Objet de la thèse

L'équipe-projet Inria Memphis est spécialisée dans la simulation de problèmes multi-physiques par des méthodes de calcul scientifique variées. Plus spécifiquement, elle s'intéresse aux écoulements de fluides visqueux incompressibles, c'est-à-dire de viscosité cinématique  $\nu > 0$  et dont la masse volumique  $\rho$  est constante. Les **équations de Navier-Stokes incompressibles** modélisent le comportement d'un tel fluide, dont l'état est décrit par les champs de vitesse  $\mathbf{u}$  et de pression  $p$ . Elles s'écrivent :

$$\begin{cases} \partial_t \mathbf{u} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u} - \nu \nabla \mathbf{u}) + \frac{1}{\rho} \nabla p = \mathbf{S}, \\ \nabla \cdot \mathbf{u} = 0, \end{cases}$$

complétées par une condition initiale et des conditions aux limites. La difficulté de résolution de ces équations est double : il faut traiter le terme non linéaire  $\mathbf{u} \otimes \mathbf{u}$  et garantir numériquement l'incompressibilité du fluide via la contrainte  $\nabla \cdot \mathbf{u} = 0$ .

## Vocabulaire et enjeux

Les équations sont résolues sur un domaine spatio-temporel  $\Omega \times [0, T]$ . En 2D, le domaine spatial est discrétisé en mailles cartésiennes  $\omega_C$  de dimensions  $\Delta x \times \Delta y$ . Le domaine temporel est quant à lui divisé en intervalles  $[t^n, t^{n+1}]$  de longueur  $\Delta t$ . Notons  $\mathbf{u}_{num}$  et  $\mathbf{u}_{ex}$  les solutions numérique et exacte du système complet, et  $\varepsilon := \|\mathbf{u}_{num} - \mathbf{u}_{ex}\|$  l'erreur commise.

**Schéma précis.** Un schéma numérique est dit **précis** respectivement à l'ordre  $p, q$  en espace et en temps lorsque  $\varepsilon = \mathcal{O}(h^p) + \mathcal{O}(\Delta t^q)$ , où  $h := \min(\Delta x, \Delta y)$ .

L'intérêt d'atteindre un ordre arbitrairement élevé  $m$  est que si l'on affine la discrétisation d'un facteur 2, l'erreur est divisée par  $2^m$ . Cette question de la précision du schéma numérique se pose d'autant plus dans le cadre de maillages de type **quadtree**, où certaines mailles peuvent être divisées (Fig. 56.1).

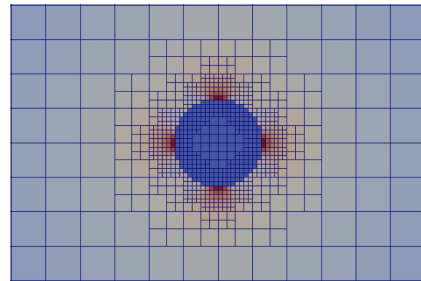


FIGURE 56.1 – Maillage quadtree autour d'un disque [1].

Ces **maillages hiérarchiques** permettent de décrire plus précisément les éléments géométriques d'intérêt dans le domaine de calcul, sans alourdir inutilement le coût de la simulation. Pourtant, aux sauts de niveau de raffinement, la précision peut se dégrader.

**Schéma compact.** Pour construire un schéma d'ordre élevé  $(d + 1)$ , on représente la solution numérique par des polynômes  $\mathbb{Q}_d$  dans chaque maille. Des schémas aux différences ou volumes finis ont été mis au point pour des quadtrees [1, 2], mais la montée en ordre n'est pas immédiate. La thèse en cours s'intéresse à des méthodes de type **Galerkin discontinu** (DG) qui permettent d'obtenir un schéma d'ordre arbitrairement élevé **compact** dans chaque maille, c'est-à-dire très peu dépendant de son voisinage.



## L'équation d'advection-diffusion

On s'intéresse ici à l'équation d'advection-diffusion, dans le cas 1D pour simplifier. Notons  $F(u)$  le flux physique, défini comme la somme de ses contributions advective  $F_a(u) := u^2$  et diffusive  $F_d(u) := -\nu \nabla u$ . L'équation s'écrit alors :  $\partial_t u + \nabla \cdot F(u) = S$ .

**Méthode ADER-DG.** Dans cette thèse, on combine des méthodes de type DG en espace à une technique ADER en temps (Arbitrary high order DERivatives). Avec cette association, l'intégration en temps est effectuée *en théorie* au même ordre de précision que celui apporté par la discrétisation DG spatiale ( $q = p$ ). Dans sa forme la plus récente [3], le schéma **ADER-DG** est alors composé de deux étapes successives (Fig. 56.2) :

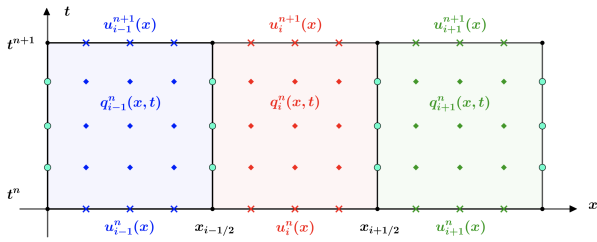


FIGURE 56.2 – Schéma ADER-DG : position des degrés de liberté espace-temps pour la prédiction et la correction ( $d = 2$ ) [5].

- **Prédiction** : on considère que chaque maille  $\omega_C$  est indépendante de ses voisines, et on détermine un prédicteur de Galerkin local  $\mathbf{q} \in \mathbb{R}^{(d+1)^2}$  afin d'obtenir une estimation de la solution :

$$\underline{A} \mathbf{q}_C^n + \underline{B} \mathbf{F}_C^n = \underline{C} \mathbf{u}_C^n + \underline{D} \mathbf{S}_C^n.$$

- **Correction** : on recombine les informations entre les mailles voisines  $\mathbf{u} \in \mathbb{R}^{(d+1)}$  et on impose les conditions aux limites à travers des flux numériques  $\tilde{\mathbf{z}} = \tilde{\mathbf{z}}(\mathbf{q}) \in \mathbb{R}^{(d+1)}$  pour obtenir :

$$\begin{aligned} \mathbf{u}_C^{n+1} &= \mathbf{u}_C^n + \underline{M} \mathbf{F}_C^n + \Delta t \underline{N} \mathbf{S}_C^n \\ &\quad - \frac{\Delta t}{\Delta x} \left( \underline{P}_R \tilde{\mathbf{F}}_{C/R}^n - \underline{P}_L \tilde{\mathbf{F}}_{L/C}^n \right) \\ &\quad + \varepsilon \frac{\Delta t}{\Delta x} \left( \underline{Q}_R \tilde{\mathbf{q}}_{C/R}^n - \underline{Q}_L \tilde{\mathbf{q}}_{L/C}^n \right), \end{aligned}$$

où  $\varepsilon \in \{-1, 0, +1\}$ , et  $\tilde{\mathbf{z}}_{A/B}$  désigne le flux numérique de la quantité  $\mathbf{z}$  entre les mailles voisines  $\omega_A$  et  $\omega_B$ .

**Ordre optimal.** Pour résoudre l'équation d'advection pure, l'utilisation de polynômes  $\mathbb{Q}_d$  produit toujours un schéma DG d'ordre de précision spatiale ( $d+1$ ). En revanche, cet ordre est limité à  $d$  pour des valeurs paires de  $d$  pour la diffusion pure [4]. Ce comportement se joue à la fois sur la valeur du paramètre  $\varepsilon$ , et sur le coefficient  $\sigma_d$  dans le flux numérique diffusif :

$$(\tilde{\mathbf{F}}_d)_{A/B} := \frac{1}{2} [F_d(\mathbf{q}_A) + F_d(\mathbf{q}_B)] + \frac{\sigma_d}{2} [\mathbf{q}_A - \mathbf{q}_B].$$

Trois versions de **pénalisation intérieure** (IP) ont été comparées en détail [4, 5] : symétrique (SIPG), incomplète (IIPG) et non symétrique (NIPG). Une technique

de **relaxation** de Cattaneo a également été étudiée. Respectivement, ces approches sont caractérisées par les paramètres :

- **SIPG** :  $\varepsilon = -1$  et  $\sigma_d := \kappa_S(d) \nu h^{-1}$ ,
- **IIPG** :  $\varepsilon = 0$  et  $\sigma_d := \kappa_I(d) \nu h^{-1}$ ,
- **NIPG** :  $\varepsilon = +1$  et  $\sigma_d := \kappa_N(d) \nu h^{-1}$ ,
- **Cattaneo** :  $\varepsilon = 0$  et  $\sigma_d := \kappa_C(d) \sqrt{\nu} h^{-\frac{d+1}{2}}$ ,

où  $\kappa(d)$  sont des coefficients à choisir. Des trois schémas IP, il a été montré numériquement que la version symétrique est toujours optimale [5]. La comparaison de la méthode de Cattaneo aux trois méthodes IP pourra donner lieu à une communication ultérieure.

## Conclusion

À ce jour, cette thèse a permis de mettre au point des méthodes de type DG jusqu'à l'ordre 5 pour les équations de Poisson et d'advection-diffusion non linéaire en 2D. Les résultats seront présentés au **CANUM 2024**, tandis que le cas scalaire et linéaire en 1D a fait l'objet d'une **publication acceptée** [5]. La suite de ce travail comporte deux volets : adapter la méthode ADER-DG aux équations de Navier-Stokes incompressibles, et considérer des maillages en quadrees. À terme, un exemple d'objectif serait de reproduire le résultat de la figure 56.3, avec une efficacité globale supérieure.

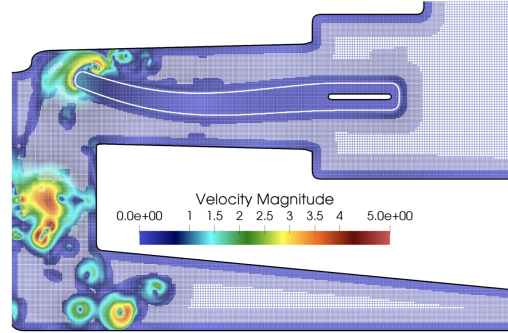


FIGURE 56.3 – Pompe cardiaque à membrane élastique. Raffinement par quadrees dans les zones de forte vitesse [2].

## Bibliographie

- [1] **Alice Raeli**, *Solution of the variable coefficients Poisson equation on Cartesian hierarchical meshes in parallel : applications to phase changing materials*, PhD, 2017.
- [2] **Antoine Fondanèche**, *Interaction fluide-structure dans un dispositif vasculaire actif*, PhD, 2021.
- [3] **M. Giuliano Carlino**, *Schéma ADER sur des maillages overset avec transmission compacte et hyper-réduction : application aux équations de Navier-Stokes incompressibles*, PhD, 2021.
- [4] **Béatrice Rivière**, *Discontinuous Galerkin methods for solving elliptic and parabolic equations : theory and implementation*, SIAM, 2008.
- [5] **Alexis Tardieu**, *High order ADER-IPDG methods for the unsteady advection-diffusion equation*, CAMC, 2024.



# EFFET RÉGULARISANTS, ANISOTROPIE ET CONTROLABILITÉ EN MÉCANIQUE DES FLUIDES

Adrien TENDANI-SOLER

Direction : Marius PAICU

Co-Direction : Sylvain ERVEDOZA

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

L'essentiel de mes travaux de recherche vise à exhiber et quantifier les effets régularisants pour des EDP issues de la mécanique des fluides. Plus précisément, je cherche à savoir si des solutions sont analytiques réelles ou holomorphes (en espace), et à quantifier cette régularité en estimant le domaine d'analyticité via différentes méthodes. Dans le travail que j'ai effectué durant ma thèse, j'ai eu l'occasion de développer cette problématique dans différents contextes, tels que l'existence de solutions analytiques ou les espaces atteignables en théorie de contrôle. En outre, j'ai travaillé sur la contrôlabilité pour un modèle d'interface diffuse pour des fluides compressibles (système de Navier-Stokes-Korteweg), ainsi que sur le caractère bien posé de certaines EDP dans différents types d'espaces fonctionnels.

## Effets régularisants et estimations de rayon d'analyticité

Chronologiquement, mon premier travail de thèse [17] portait sur des estimations de rayon d'analyticité pour le système de Navier-Stokes-Korteweg (NSK) (je renvoie à [18] pour une présentation du système), dans des espaces de pseudo-mesures homogènes. J'ai ensuite exploré cette problématique dans le cadre du système de Navier-Stokes sous-riemannien (NSsR) 57.1 sur  $\mathbb{R}^3$ , où j'ai montré que l'on avait une meilleure estimation (en temps suffisamment grand) de la bande d'analyticité en temps grands (pour des solutions globales) en variable verticale comparée à ce qui a été démontré dans le cas du système de Navier-Stokes classique sur  $\mathbb{R}^3$ .

Pour un espace de Banach  $E$  qui se plonge dans  $\mathcal{S}'(\mathbb{R}^d)$ , on définit le rayon d'analyticité

$$\text{rad}_E(u) := \sup \left\{ \sigma \geq 0 \mid e^{\sigma|D_x|} u \in E \right\}.$$

Le lien avec la bande d'analyticité est le suivant : si  $\text{rad}_E(u) > \sigma > 0$ , alors on s'attend à ce que la fonction  $u$  se prolonge en une fonction holomorphe sur la bande  $\{z \in \mathbb{C}^d \mid |\Im(z)| < \sigma\}$  (voir l'appendice de [17] pour cette interprétation dans les espaces de pseudo-mesures et le théorème 1.1 de [2] dans les espaces de Sobolov). Que ça soit pour le système de Navier-Stokes incompressible ou (NSK) sur  $\mathbb{R}^d$  le résultat typique est

$$\text{rad}_E(u(t)) \geq c\sqrt{t},$$

où  $E$  est un certain espace de Banach. Dans [11] et dans [4] des améliorations ont été apportées pour le système de Navier-Stokes incompressible près de 0 dans les espaces surcritiques et en temps grand dans les espaces critiques.

## Amélioration du rayon d'analyticité pour le système de Navier-Stokes-Korteweg

Mon objectif dans [17] était d'améliorer l'estimation du rayon d'analyticité près de  $t = 0$  obtenue par Charve, Danchin et Xu dans [3] en le comparant à  $\sqrt{t|\ln(t)|}$  au lieu de  $\sqrt{t}$ , à la fois pour le champ des vitesses et la densité, en utilisant une méthode récente développée pour une classe d'EDP paraboliques semilinéaires dans [4] par Chemin, Gallagher et Zhang qui ont aussi montré un résultat similaire à celui trouvé par Herbst et Skibsted pour l'équation de Navier-Stokes sur  $\mathbb{R}^3$  dans [11]. L'amélioration de la bande d'analyticité que j'ai obtenue concerne uniquement les solutions locales dans des espaces sur-critiques. Dans [17], j'obtiens en plus l'existence, globale dans un espace critique et locale dans des espaces sur-critiques, avec dans chaque cas, l'analyticité de la solution, et l'unicité. Le cadre fonctionnel développé dans cet article vise à obtenir ces résultats dans des espaces où le contrôle de la norme de la solution permet d'estimer la décroissance ponctuelle en temps et en fréquence.

## Théorie de Cauchy et effet regularisant pour le système de Navier-Stokes sous-Riemannien

Ce système fait intervenir deux inconnues  $(u, p)$  où  $u : \mathbb{R}_+ \times \mathbb{R}^3 \rightarrow \mathbb{R}^2$  est le champ des vitesses (qui est anisotrope) et  $p : \mathbb{R}_+ \times \mathbb{R}^3 \rightarrow \mathbb{R}$  la pression, et s'écrit

$$\begin{cases} \partial_t u_1 + u_1 X u_1 + u_2 \Xi u_1 - \Delta_{\mathbb{H}} u_1 = -X p \\ \partial_t u_2 + u_1 X u_2 + u_2 \Xi u_2 - \Delta_{\mathbb{H}} u_2 = -\Xi p \\ X u_1 + \Xi u_2 = 0, \end{cases} \quad (57.1)$$

où  $u = {}^t(u_1, u_2)$ ,  $X := \partial_x + 2y\partial_z$ ,  $\Xi := \partial_y - 2x\partial_z$  et  $\Delta_{\mathbb{H}} := X^2 + \Xi^2$ , avec  $(x, y, z) \in \mathbb{R}^3$ . Je propose dans [20] une dérivation formelle de ce modèle, où le champ des vitesses admet une anisotropie qui est donnée par une structure géométrique : au lieu de prendre

$u : \mathbb{R}^3 \rightarrow T\mathbb{R}^3$ , on choisit  $u : \mathbb{R}^3 \rightarrow H \subset T\mathbb{R}^3$ , avec  $H$  un sous-fibré de  $T\mathbb{R}^3$  définissant une structure sous-Riemannienne sur  $\mathbb{R}^3$  (voir [1], Définition 3.2, Page 68). Cette structure sous-Riemannienne est liée à un groupe de Lie appelé groupe de Heisenberg. Je renvoie à l'excellente introduction de [1] pour une présentation informelle de cette structure.

**1) Solution d'énergie finie, caractère bien posé et effets régularisants** Dans [20], j'étudie l'existence globale de solutions faibles pour des données dans  $L^2$  et ensuite le caractère bien posé dans l'espace critique  $\tilde{H}^1$  associé à une certaine famille de champ de vecteurs  $\tilde{X}$  et  $\tilde{\Xi}$ . Le résultat qui apporte quantitativement une nouveauté par rapport au système de Navier-Stokes<sup>13</sup> sur  $\mathbb{R}^3$  concerne la bande d'analyticité en la variable verticale  $z$  de la solution. En effet, on obtient pour tout  $t > 0$

$$\text{rad}(u(t)) \geq ct,$$

où  $c \in (0, 4)$  et  $\text{rad}$  est le rayon d'analyticité en la variable  $z$  uniquement. Puis finalement je montre que la solution est (seulement) lisse par rapport à toutes les variables d'espace et de temps. Notons que l'on ne s'attend pas à avoir de l'analyticité dans les variables d'espace  $x$  et  $y$ , à cause du manque d'ellipticité de  $\Delta_H$  (qui est néanmoins hypoelliptique).

**2) Caractère bien posé dans des espaces de régularité anisotrope** Dans ce deuxième travail sur (NSsR), je m'intéresse à l'existence globale et à l'unicité des solutions dans des espaces ne faisant intervenir que la régularité par rapport à la variable " $z$ ". Plus précisément, je montre l'existence et l'unicité des solutions (analytiques) associées à des données initiales  $u_0$  telles qu'au moins une des quantités suivantes

$$\|u_0\|_{L^2} \|\partial_z u_0\|_{L^2} \quad \text{ou} \quad \|u_0\|_{B^{0,1/2}},$$

soient petites. Notons que  $B^{0,1/2}$  correspond, en terme de régularité, à une demi-dérivée verticale (voir [15]). L'introduction dans [19] de nouvelles inégalités de type Gagliardo-Nirenberg stratifiées sur le groupe d'Heisenberg permet d'adapter les stratégies utilisées pour le système de Navier-Stokes anisotrope sur  $\mathbb{R}^3$ , notamment celle de [15]. Je donne aussi une estimation de la bande d'analyticité dans  $B^{0,1/2}$  similaire à celle montrée dans [20].

## Contrôlabilité et espace atteignable

### Contrôlabilité du système de Navier-Stokes Korteveg

J'ai étudié la contrôlabilité exacte locale vers des trajectoires constantes pour (NSK) sur le tore en dimension  $d \in \{1, 2, 3\}$  lorsque le contrôle agit sur un sous-ensemble ouvert. La contrôlabilité interne sur le tore avec des contrôles agissants sur toutes les équations est un exemple plutôt académique. L'intérêt (en vue d'application physique) de ce résultat est qu'il a pour corollaire un résultat de contrôlabilité exacte dans des domaines bornés réguliers  $\Omega$  lorsque le contrôle agit

sur toute la frontière  $\partial\Omega$  et toujours vers les mêmes trajectoires  $(\rho_*, 0)$ .

Tout d'abord, à ma connaissance ce résultat de contrôlabilité est le premier concernant (NSK). En outre, ce résultat est valable quel que soit  $\omega$  et pour tout temps final  $T > 0$  et il en est de même pour le système linéarisé de (NSK). Ceci contraste fortement avec les résultats de contrôlabilité pour l'équation de Navier-Stokes compressible (correspondant au cas  $\kappa = 0$ ) linéarisée. En effet, pour ce dernier, les résultats de contrôlabilité exacte locale autour des trajectoires  $(\rho_*, u_*)$ , sont valides sous réserve de conditions géométriques appropriées liées au flux du champ de vitesse cible, pour  $u_* \neq 0$  et  $T$  suffisamment grand, (voir par exemple [6], [7] et [8]).

### Espace atteignable pour le système de Stokes

Je présente dans cette dernière section mes travaux en collaboration avec Sylvain Ervedoza sur l'espace atteignable pour le système de Stokes.

**Sur l'espace atteignable :** La problématique générale des espaces atteignables peut être résumée de la manière suivante : *Pour un système contrôlé donné. Étant donné un état initial  $u_i$  et un temps  $T > 0$ , décrire l'espace  $R(u_i, T)$  des états finaux  $u_f$  que l'on peut atteindre à partir de  $u_i$  au temps  $T$ .* Donner une caractérisation précise des états qui peuvent être atteints en un certain temps  $T > 0$  est une question encore largement ouverte : même pour l'équation de la chaleur à coefficients constants en une dimension, contrôlée depuis la frontière, la caractérisation complète de l'espace atteignable, en termes d'espaces de Bergman, n'a été obtenue que très récemment dans [12]. La description de l'espace atteignable pour l'équation de la chaleur a reçu une grande attention ces dernières années (voir [13, 5, 9, 12, 14, 10] pour le cas de la dimension 1 et [16] pour le cas multidimensionnel).

Soit  $\Omega$  une boule de  $\mathbb{R}^d$  avec  $d \in \{2, 3\}$ . En notant  $\mathcal{R}$  l'espace atteignable à partir de l'état nul pour le problème de Stokes, j'obtiens alors le résultat suivant

$$\mathcal{O}_{div}(\overline{\mathcal{E}_\Omega}) \subset \mathcal{R} \subset \mathcal{O}_{div}(\mathcal{E}_\Omega),$$

où  $\mathcal{O}_{div}(U)$  est l'espace des fonctions holomorphes sur  $U$  à divergence nulle et  $\mathcal{E}_\Omega$  l'ensemble des  $z \in \mathbb{C}^d$  tels que  $\Re(z) \in \Omega$  et  $|\Im(z)| < \text{dist}(\Re(z), \partial\Omega)$ . Ce résultat est similaire à ce qui a été montré pour l'équation de la chaleur dans [16].

## Bibliographie

- [1] A. Agrachev, D. Barilari, and U. Boscain. [A Comprehensive Introduction to Sub-Riemannian Geometry](#). Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2019.
- [2] T. Alazard, N. Burq, and C. Zuily. Cauchy theory for the water waves system in an analytic framework. *Tokyo Journal of Mathematics*, 45(1) :103 – 199, 2022.
- [3] F. Charve, R. Danchin, and J. Xu. Gevrey analyticity and decay for the compressible Navier-

13. Pour lequel le meilleur score en temps long est essentiellement  $\liminf_{t \rightarrow +\infty} \text{rad}(u(t))/\sqrt{t} > 0$  (voir [4]).

- Stokes system with capillarity. Indiana University Mathematics Journal, 2018.
- [4] J.-Y. Chemin, I. Gallagher, and P. Zhang. On the radius of analyticity of solution to semi-linear parabolic systems. Math Research Letters, 8, 2020.
  - [5] J. Darde and S. Ervedoza. On the reachable set for the one-dimensional heat equation. SIAM Journal on Control and Optimization, 56 :1692–1715, 2018.
  - [6] S. Ervedoza, O. Glass, and S. Guerrero. Local exact controllability for the two- and three-dimensional compressible Navier-Stokes equations. Communications in Partial Differential Equations, 41 :1660 – 1691, 2015.
  - [7] S. Ervedoza, O. Glass, S. Guerrero, and J.-P. Puel. Local exact controllability for the one-dimensional compressible Navier-Stokes equation. Arch. Ration. Mech. Anal., 206(1) :189–238, 2012.
  - [8] S. Ervedoza and M. Savel. Local boundary controllability to trajectories for the 1D compressible Navier Stokes equations. ESAIM Control Optim. Calc. Var., 24(1) :211–235, 2018.
  - [9] A. Hartmann, K. Kellay, and M. Tucsnak. From the reachable space of the heat equation to hilbert spaces of holomorphic functions. Journal of the European Mathematical Society, 22 :3417–3440, 2020.
  - [10] A. Hartmann and M.-A. Orsoni. Separation of singularities for the bergman space and application to control theory. Journal de Mathématiques Pures et Appliquées, 150 :181–201, 2021.
  - [11] I. Herbst and E. Skibsted. Analyticity estimates for the Navier-Stokes equations. Advances in Mathematics, 228(4) :1990–2033, 2011.
  - [12] K. Kellay, T. Normand, and M. Tucsnak. Sharp reachability results for the heat equation in one space dimension. Analysis and PDE, 15(4) :891 – 920, 2022.
  - [13] P. Martin, L. Rosier, and P. Rouchon. On the reachable states for the boundary control of the heat equation. Applied Mathematics Research eXpress, 2016(2) :181–216, 2016.
  - [14] M.-A. Orsoni. Reachable states and holomorphic function spaces for the 1-d heat equation. Journal of Functional Analysis, 280 :108852, 2021.
  - [15] M. Paicu. Équation anisotrope de Navier-Stokes dans des espaces critiques. Revista Matemática Iberoamericana, 21 :179–235, 2005.
  - [16] A. Strohmaier and A. Waters. Analytic properties of heat equation solutions and reachable sets. Mathematische Zeitschrift, 302(1) :259–274, 2022.
  - [17] A. Tendani-Soler. Analytic regularity for Navier-Stokes-Korteweg model on pseudo-measure spaces. Dynamics of Partial Differential Equations, 20(1) :1–21, 2023.
  - [18] A. Tendani-Soler. Local exact controllability to constant trajectories for the navier-stokes-korteweg model. Submitted, Preprint in ArXiv, 2023.
  - [19] A. Tendani-Soler. Anisotropic framework for sub-riemannian navier-stokes system. En cour de finalisation.
  - [20] A. Tendani-Soler. Sub-riemannian navier-stokes system : well-posedness and regularity. En cour de finalisation.

# MULTIVARIATE QUANTILES, OPTIMAL TRANSPORT

Gauthier THURIN

Advisor : Jérémie BIGOT

Co-advisor : Bernard BERCU

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

**M**onge-Kantorovich quantiles are defined in an intuitive fashion through the optimal transportation from a reference distribution. My thesis is dedicated to their study, as they benefit from most of sought-after properties. The adaptivity of these concepts makes them attractive for building statistical tools, and has concentrated much attention in the recent years. My PhD project partly belongs to this line of work, as we proposed multivariate definitions of superquantiles, expected shortfalls and related risk measures. In addition, we focused on the estimation of a regularized version of the Monge-Kantorovich quantile function, via a stochastic algorithm with almost-sure convergence guarantees.

## Ordering multivariate data

Consider a probability distribution  $\nu$  supported on  $\mathbb{R}^d$ . In the scalar case  $d = 1$ , the quantile function of  $\nu$  is nothing else than the generalized inverse  $F_\nu^{-1}$  of the cumulative distribution function  $F_\nu$ . This heavily relies on the left-to-right ranking of samples from  $\nu$ . However, in the multi-dimensional case  $d > 1$ , the lack of a canonical ordering precludes a consensual notion of quantiles, see [1, 2] and references therein. The key ingredient for a concept of quantiles is thus the ordering it provides for observations sampled from  $\nu$ .

Departing from what is known,  $F_\nu^{-1}$  appears to be the unique solution of the optimal transport (OT) problem between the uniform  $\mu = U([0, 1])$  and  $\nu$ . By stating this as a definition when  $d > 1$ , the authors of [1] generalized univariate quantiles to define the Monge-Kantorovich (MK) quantile function of  $\nu$  as

$$\mathbf{Q} = \operatorname{argmin}_{T: T_\# \mu = \nu} \mathbb{E}_{X \sim \mu} (\|X - T(X)\|^2),$$

where the push-forward constraint  $T_\# \mu = \nu$  means

$$\int_{\mathcal{X}} g d\nu = \int_{T^{-1}(x)} g \circ T d\mu.$$

A suitable choice of reference measure  $\mu$  is the one of the random vector  $R\Phi$ , where  $R$  and  $\Phi$  are independent and drawn uniformly from  $[0, 1]$  and the unit hypersphere  $\mathbb{S}^{d-1} = \{\varphi \in \mathbb{R}^d : \|\varphi\| = 1\}$ , respectively. With this choice, the unit balls  $\mathbb{B}(0, \alpha)$  have  $\mu$ -probability  $\alpha$ , and the regions  $\mathbf{Q}(\mathbb{B}(0, \alpha))$  are nested and have  $\nu$ -probability  $\alpha$ , and thus are appropriate candidates for quantile regions, [1, 2]. Descriptive plots in the spirit of multivariate boxplots are given in Figure 58.1 for a banana-shaped distribution  $\nu$ . The fact that the random vector  $\mathbf{Q}(U)$  follows  $\nu$  as soon as  $U$  is sampled from  $\mu$  is crucial. It allows to claim that  $\mathbf{Q}$  is able to capture all the available information. Roughly, in Figure 58.1, the same amount of mass is contained in the regions

delimited by two red curves, and the blue curves encapsulate quantile regions indexed by their probability contents.

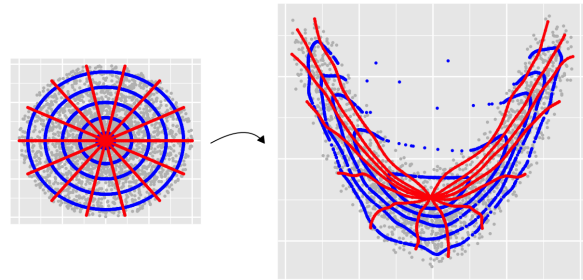


FIGURE 58.1 – For any distribution, on the right, a transported ordering from some reference measure, on the left.

## Statistical tools

In  $\mathbb{R}$ , the superquantiles and expected shortfalls complement the information given by the quantiles. By focusing on which univariate features we aim to extend, we defined in [3] the MK superquantile function

$$S(u) = \frac{1}{1 - \|u\|} \int_{\|u\|}^1 \mathbf{Q}(t \frac{u}{\|u\|}) dt,$$

and the MK expected shortfall function

$$E(u) = \frac{1}{\|u\|} \int_0^{\|u\|} \mathbf{Q}(t \frac{u}{\|u\|}) dt.$$

The counterpart of Figure 58.1 by using  $S$  and  $E$  instead of  $\mathbf{Q}$  induces the descriptive plots of Figure 58.2. These functions describe central and peripheral areas of point clouds, and they can be shown to characterize random vectors and their convergence in distribution, [3].

**Theorem.** Let  $\nu_1$  and  $\nu_2$  be probability distributions on  $\mathbb{R}^d$ , with respective MK superquantile and expected shortfall functions given by  $S_1, E_1$  and  $S_2, E_2$ . Then,

$$\nu_1 = \nu_2 \Leftrightarrow S_1 = S_2 \text{ } \mu\text{-a.e.} \Leftrightarrow E_1 = E_2 \text{ } \mu\text{-a.e.}$$

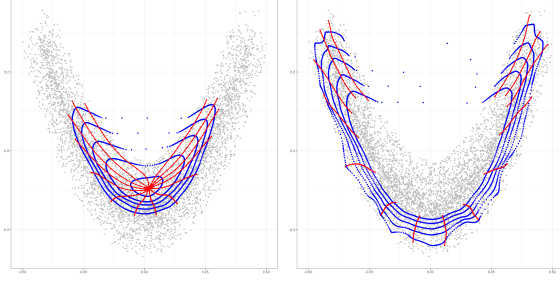


FIGURE 58.2 – Descriptive plots,  $E$  on the left,  $S$  on the right.

**Theorem.** Let  $\nu_n$  and  $\nu$  be such that their respective MK quantile functions  $Q_n$  and  $Q$  are continuous on  $\mathbb{B}(0, 1) \setminus 0$ . Then the weak convergence of  $\nu_n$  towards  $\nu$  is equivalent to

$$\forall u \in \mathbb{B}(0, 1) \setminus 0, \lim_{n \rightarrow +\infty} E_n(u) = E(u)$$

and, for any compact  $K \subset \mathbb{B}(0, 1) \setminus 0$ ,

$$\lim_{n \rightarrow +\infty} \sup_{u \in K} \|E_n(u) - E(u)\| = 0.$$

Besides, if one assumes that, for every  $n \in \mathbb{N}$  and  $X_n$  drawn from  $\nu_n$ , there exists a random vector  $Z$  with  $\mathbb{E}[Z \ln(Z)] < +\infty$  such that  $\|X_n\| \leq Z$ , then the above weak convergence is also equivalent to

$$\forall u \in \mathbb{B}(0, 1) \setminus 0, \lim_{n \rightarrow +\infty} S_n(u) = S(u).$$

Perhaps the most classical univariate application of these notions is the measurement of risk, with the fundamental risk measures given by the Value-at-Risk (VaR) and the Conditional-Value-at-Risk (CVaR). The risk framework considers a vector of losses  $X \in \mathbb{R}_+^d$  where each component is a positive measure. From our definitions, one can just select a point with maximal norm within a MK quantile contour (resp. superquantile) to get a multivariate VaR (resp. CVaR), see [3] for further detail. Such risk measurements summarize a given dataset to answer the following :

- With probability  $\alpha$ , what is the worst that can happen ?
- In case the worst happens, what shall we expect, in average ?

Using MK quantiles allows to take into account the multivariate probability for the assertion “with probability  $\alpha$ ”. These are illustrated in Figure 58.3, with different choices of reference measure that correspond to a center-outward or a left-to-right ordering for  $\nu$ .

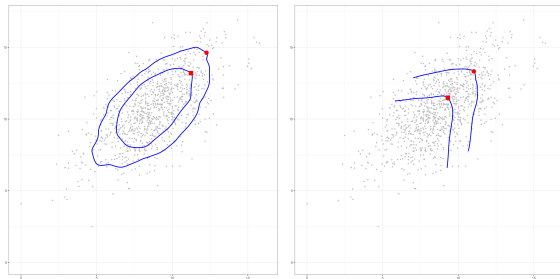


FIGURE 58.3 – VaR and CVaR, center-outward or left-to-right ordering.

## Estimation

In practice, the estimation of  $Q$  amounts to solve an OT problem involving the empirical measure associated to samples  $Y_1, \dots, Y_n$  from  $\nu$ . Such problem yields a quantile function whose values are constrained to belong to the  $(Y_i)$ , unless one uses regularization in a second step. Interpolation is desirable as soon as the focus is on quantile contours and regions, or if it is required to compute out-of-sample estimates  $Q(x)$ , for any  $x$ . The estimation of regularized MK quantiles is another direction of my thesis, in which we advocated the use of the entropic regularization of optimal transport, [4]. This results in the *entropic map*, that has the desirable feature of being the gradient of a convex function, even in practice, and that is rooted in the literature on computational optimal transport.

Therefore, the estimation procedure requires to solve the entropically regularized OT problem. Making use of the fact that one of the two distributions is held fixed, we were able to design a new algorithm tackling the continuous OT in the limit  $n \rightarrow +\infty$ , for  $n$  the size of the sample  $(Y_i)$ . The idea is to parametrize dual potentials in the Kantorovich formulation of OT by their Fourier coefficients, leading to a stochastic gradient descent on absolutely summable sequences. Using stochastic algorithms has two major consequences. On the practical side, it allows to avoid the storage of the cost matrix of size  $n^2$  between two samples of size  $n$ . On the theoretical side, one can use tools from random processes in order to obtain consistency results, as in [4] where we showed the almost sure convergence of the iterates of our algorithm.

## Bibliographie

- [1] Chernozhukov, M., Galichon, A., Hallin, M., Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1), 223-256.
- [2] Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension  $d$  : A measure transportation approach. *The Annals of Statistics*, 49(2), 1139-1165.
- [3] Bercu, B., Bigot, J., and Thurin, G. (2023). Monge-Kantorovich superquantiles and expected short-falls with applications to multivariate risk measurements. *arXiv preprint arXiv :2307.01584*.
- [4] Bercu, B., Bigot, J., and Thurin, G. (2023). Stochastic optimal transport in Banach Spaces for regularized estimation of multivariate quantiles. *arXiv preprint arXiv :2302.00982*.



# MÉTHODES NUMÉRIQUES CINÉTIQUES NOVATRICES POUR LA RADIOTHÉRAPIE ET LES PLASMAS

Thomas VIGIER

Direction : Luc MIEUSSENS

Co-Direction : Julien MATHIAUD

Encadrement : Jean-Luc FEUGEAS

CELIA, Univ. Bordeaux, CNRS, CEA, UMR 5107, F-33400 Talence, France

**L**es modèles aux moments avec des fermetures adaptées peuvent donner lieu à des solveurs précis et efficace dans la résolution d'équations cinétiques. Dans cette thèse, nous proposons une nouvelle approche pour construire des schémas préservant l'asymptotique (AP), dans un premier temps pour le modèle M1 du transport linéaire puis pour le modèle M1 du transport thermique non local. Dans les deux cas, l'idée principale repose sur l'application de la fermeture M1 à l'échelle numérique dans un schéma AP préexistant pour l'équation cinétique correspondante. Ce schéma est l'*Unified Gas Kinetic Scheme* (UGKS) présenté originalement dans [1] et étendu pour le transport linéaire dans [2]. La construction et les propriétés de ces différents schémas sont étudiées.

## Le modèle M1 du transport linéaire

L'équation du transport linéaire décrit l'évolution de la fonction de distribution  $f$  d'une population de particules dans l'espace des phases  $\mathbb{R} \times [-1, 1]$  au cours du temps. Celle-ci traduit deux phénomènes physiques différents, l'advection (transport) des particules ainsi que les interactions avec le milieu. En une dimension d'espace cette équation prend la forme suivante :

$$\partial_t f + \frac{v}{\epsilon} \partial_x f = \frac{\sigma}{\epsilon^2} (\rho - f). \quad (59.1)$$

avec  $\rho = \int_{-1}^1 f(v) dv$  la densité de particules,  $v \in [-1, 1]$  la vitesse des particules,  $\sigma$  l'opacité du milieu et  $\epsilon$  le nombre de Knudsen, défini comme étant le rapport entre le libre parcours moyen des particules et une longueur de référence. Résoudre cette équation est coûteux d'un point de vue numérique à cause de la dimension du problème. Pour ce faire un modèle intégré en direction, dit "aux moments" peut être résolu à la place en supposant la forme de la fonction de distribution en vitesse. Le modèle M1 est la première hiérarchie qui permet de restaurer des phénomènes cinétiques :

$$\begin{cases} \partial_t \rho + \frac{1}{\epsilon} j = 0 \\ \partial_t j + \frac{1}{\epsilon} \rho \left( 1 - 2 \frac{j/\rho}{g(j/\rho)} \right) = 0 \end{cases} \quad (59.2)$$

Le fermeture M1 a été utilisée pour exprimer le dernier flux à partir des inconnus du système que sont la densité  $\rho$  et la densité de courant  $j = \int_{-1}^1 v f(v) dv$ . Cette repose sur un argument entropique pour imposer une forme particulière à la fonction de distribution. Celle-ci minimise l'entropie de Boltzmann sous la contrainte de la réalisation des moments et est  $\forall v \in [-1, 1], \hat{f}(v) = e^{\alpha + \beta v}$ . Élaborer un schéma nu-

mérique pour l'équation cinétique ou le modèle au moment est complexe du fait des régimes asymptotiques. En effet, lorsque le nombre de Knudsen tend vers 0 les collisions prédominent devant le transport et un phénomène de diffusion peut être observé à l'échelle macroscopique. Par ailleurs, à l'échelle microscopique le terme source devient raide et difficile à traiter numériquement. Au contraire lorsque le nombre de Knudsen est grand les particules sont majoritairement advectées. Pour correctement résoudre chacun des deux régimes, il est nécessaire de développer des schémas numériques qui préservent l'asymptotique (AP).

Dans cette thèse, nous proposons une nouvelle méthode pour obtenir des schémas AP pour des modèles aux moments à partir d'un schéma numérique AP pré-existant pour l'équation cinétique. Ce schéma est UGKS (*Unified Gas Kinetic Scheme*) et repose sur la solution intégrale de l'équation cinétique pour élaborer les flux. Cette méthode présentent l'avantage de prendre en compte le terme source dans les flux et ainsi de naturellement obtenir un bon schéma. Il est alors possible de montrer formellement que le schéma discrétise bien l'équation dans chacun des régimes. L'idée principale pour obtenir un schéma pour le modèle aux moments M1 est d'appliquer la fermeture entropique à l'échelle numérique dans les flux d'UGKS. Cette procédure s'apparente à une projection dans l'ensemble des fonctions de distribution M1 à chaque pas de temps dans UGKS.

Plusieurs extensions à ce schéma sont proposées. Une montée à l'ordre 2 n'altérant pas les régimes asymptotiques est développée. De plus le caractère général de la méthode est démontré en étendant ce travail au modèle P1 et au modèle M2. Une extension permettant d'assurer la réalisabilité des moments est aussi proposée. Finalement ce nouveau schéma est validé sur différents cas tests.

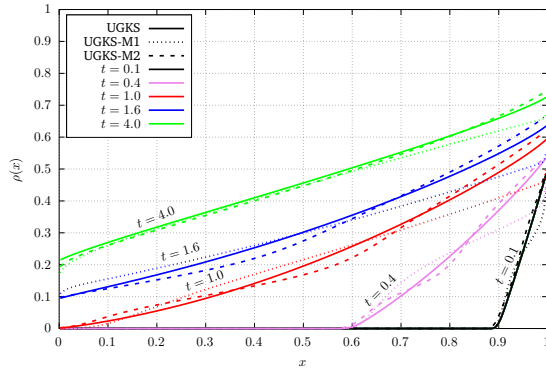


FIGURE 59.1 – Densité dans le domaine à différents instants pour UGKS, UGKS-M1 et UGKS-M2

Sur la figure 59.1, la densité dans le domaine est représentée à différents instants pour UGKS et le nouveau schéma pour les modèles M1 et M2. On constate que l'erreur commise par le modèle est réduite en augmentant la taille de la hiérarchie du modèle.

### Le modèle M1 du transport thermique non local

Ce travail dans le cadre du transport linéaire a pu être étendu dans un contexte physique plus pertinent et proche des applications. Dans les codes hydrodynamiques pour la fusion par confinement inertiel, le flux de chaleur doit être fermé. Pour ce faire une loi dite locale peut être utilisée comme celle de Spitzer-Härm. Néanmoins dans de nombreux cas hors équilibre, celle-ci n'est pas suffisante et une résolution cinétique du transport électronique est nécessaire [3]. L'équation qui décrit ce phénomène est

$$\partial_t f + \frac{v}{\epsilon} \cdot \nabla_x f = \frac{\sigma(T)}{\epsilon^2} (M[f] - f), \quad (59.3)$$

où dans ce cas la fonction de distribution est à valeur dans  $\mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{S}^2$ ,  $\Omega$  est un vecteur sur la sphère,  $\sigma$  est l'opacité dépendant de la température  $T$  et  $M[f]$  est l'équilibre Maxwellien. Pour les cas d'application, la résolution d'un modèle aux moments angulaire est suffisante pour restituer les effets cinétiques. Contrairement au modèle 1D, une variables cinétique est conservée dans la hiérarchie de moments. La même démarche est alors suivie pour obtenir un schéma numérique pour le modèle M1 associé à cette équation. De même une extension à l'ordre 2 est proposée.

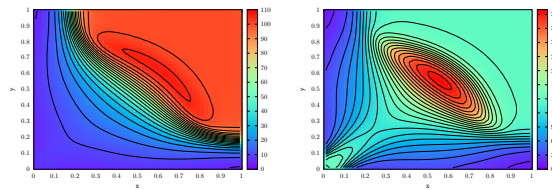


FIGURE 59.2 – Densité (à droite) et énergie (à gauche) dans le domaine dans le cas d'un dépôt d'énergie dans une zone à forte densité (avec UGKS-M1 d'ordre 2)

Sur la figure 59.2, la densité et l'énergie sont représentées dans le cas d'un dépôt d'énergie dans une zone à

forte densité. Ce cas test met en évidence les capacités du schéma à capturer les différents régimes.

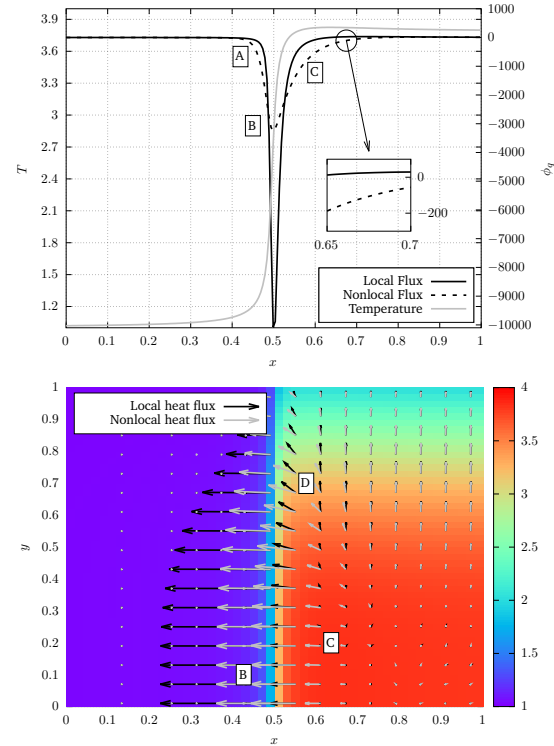


FIGURE 59.3 – Température et flux de chaleur local et non-local dans le domaine

Sur la figure 59.3, des phénomènes caractéristiques du transport thermique non local sont mis en évidence sur le flux de chaleur dans le cas de la relaxation d'un gradient de température en 1D et en 2D. On peut en particulier observer le pré-chauffage (A), la limitation du flux de chaleur (B), le flux anti-naturel (C) et enfin la rotation du flux (D).

### Bibliographie

- [1] A unified gas-kinetic scheme for continuum and rarefied flows, Xu, Kun and Huang, Juan-Chen, Journal of Computational Physics, 229, 20, 7747–7764, 2010, Elsevier
- [2] On the Asymptotic Preserving property of the Unified Gas Kinetic Scheme for the diffusion limit of linear kinetic models, Mieussens, Luc, Journal of Computational Physics, 253, 138–156, 2013, Elsevier
- [3] A practical nonlocal model for heat transport in magnetized laser plasmas, Nicolaï, Ph D and Feugeas, J-LA and Schurtz, GP, Physics of plasmas, 13, 3, 2006, AIP Publishing

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 : 2018-2021
- Guillaume Blin : 2022-