

Journées SMAI MODE

Lyon

27-29 mars 2024



Informations diverses

- **Mini cours (25-26 mars 2024)**

Salle 435, ENS de Lyon site Monod, 46 allée d'Italie, 69007 Lyon

Transport en commun : Métro B arrêt Debourg, Tram T1 arrêt ENS de Lyon

- **Journées SMAI-MODE (27-29 mars 2024)**

INSA Lyon, Bâtiment Gustave Ferrié, 8 rue de la Physique, 69100 Villeurbanne

Transport en commun : Tram T1, Tram T4 arrêt La Doua - Gaston Berger

- **Conférence grand public (28 mars 2024, 17h45)**

Amphi Jordan, Bâtiment Braconnier, 21 Avenue Claude Bernard, 69100 Villeurbanne

Transport en commun : Tram T1, Tram T4 arrêt Université Lyon 1

- **Dîner de Gala (28 mars 2024, 20h)**

Bocuse l'Ouest, 1 Quai du Commerce, 69009 Lyon

Transport en commun : Métro D

- **Repas 27-29 mars 2024 midi**

Table d'Einstein, Délégation régionale, 2 av. Albert Einstein, 69100 Villeurbanne

Programme mercredi 27 mars 2024 (matin)

9h00 - 9h15	Mots de bienvenue Amphi AE1
9h15 - 10h15	Conférence plénière Amphi AE1 – Session Chair: Filippo Santambrogio Anne Auger , Inria <i>On the linear convergence of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)</i>
10h15-10h30	Hommage à Hedy Attouch par Samir Adly Amphi AE1
11h00 - 13h00	Méthodes du premier ordre et leurs applications à l'apprentissage Amphi AE1 – Session chair: Adrien Taylor
11h00 - 11h30	Waïss Azizian , LJK, UGA <i>The interplay between geometry and convergence in Bregman proximal methods</i>
11h30 - 12h00	Mathieu Even , Inria, ENS <i>Asynchronous Speedup in Centralized and Decentralized Optimization</i>
12h00 - 12h30	Guillaume Olikier , Université Côte d'Azur, Inria <i>Projected gradient descent accumulates at Bouligand stationary points</i>
12h30 - 13h00	Scott Pesme , EPFL <i>Saddle-to-Saddle Dynamics in Diagonal Linear Networks</i>
11h00 - 13h00	Transport Salle E010 – Session chair: Yohann De Castro
11h00 - 11h30	Pierre-Cyril Aubin , TU Wien <i>Extending convexity and gradient descent: a framework for general costs based on alternating minimization</i>
11h30 - 12h00	Tam Le , Université Grenoble-Alpes <i>Nonsmoothness in Wasserstein distributionally robust models</i>
12h00 - 12h30	Andrea Natale , Inria, Univ. Lille <i>A second-order accurate space-time discretization for L2-Wasserstein gradient flows with exact energy dissipation</i>
12h30 - 13h00	Gabriele Todeschi , Université Gustave Eiffel <i>Extrapolation dans l'espace de Wasserstein</i>
11h00 - 13h00	Contrôle optimal Salle E011 – Session chair: Florent Nacry
11h00 - 11h30	Arthur Bottois , L2S CentraleSupélec, Inria Saclay-Île-de-France <i>Analyse de sensibilité et contrôle optimal pour la Chemical Master Equation</i>
11h30 - 12h00	Anas Bouali , Avignon University <i>Loss control regions in optimal control problems</i>
12h00 - 12h30	Rémy Dutto , Vitesco, IRIT, IMT <i>Préconditionnement géométrique de la méthode indirecte de tir appliquée au véhicule hybride</i>
12h30 - 13h00	Piero Visconti , INSA Rouen Normandie <i>Optimality conditions for parabolic stochastic optimal control problems with boundary controls</i>

Programme mercredi 27 mars 2024 (après-midi)

14h30 - 15h30	Conférence plénière
	Amphi AE1 – Session chair: Nelly Pustelnik
	Emilie Chouzenoux , Inria Saclay <i>Deep Unfolding Approach for Limited-Angle Computed Tomography Image Reconstruction</i>
16h00 - 18h00	Théorie des jeux
	Salle E010 – Session chair: Guillaume Vigeral
16h00 - 16h30	Côme Fiegel , ENSAE CREST <i>Adapting to game trees in zero-sum imperfect information games</i>
16h30 - 17h00	Ivan Novikov , Université Paris-Dauphine <i>Zero-Sum Stochastic Games with Vanishing Stage Duration and Public Signals</i>
17h00 - 17h30	Vianney Perchet , CNRS ENSAE CREST <i>Trading-off price for data quality to achieve fair online allocation</i>
17h30 - 18h00	Bruno Ziliotto , CEREMADE, CNRS, Université Paris Dauphine <i>Le jeu derrière la percolation orientée</i>
16h00 - 18h00	Optimisation
	Amphi AE2 – Session chair: Samir Adly
16h00 - 16h30	Assalé Adjé , LAMPS - Université de Perpignan Via Domitia <i>Maximum of Reachable Values of Discrete-Time Systems</i>
16h30 - 17h00	Xavier Dupuis , Institut de Mathématiques de Bourgogne <i>Le chemin des solutions de l'estimateur SLOPE</i>
17h00 - 17h30	Florentin Goyens , Paris Dauphine - PSL <i>Strict saddle optimization on manifolds</i>
17h30 - 18h00	Sorin-Mihai Grad , UMA/ENSTA Paris <i>Beyond the Fermat optimality rules</i>
16h00 - 18h00	Optimisation et contrôle optimal dans l'espace de Wasserstein
	Amphi AE1 – Session chair: Othmane Jerhaoui
16h00 - 16h30	Giulia Cavagnari , Politecnico di Milano <i>Optimal control problems for multi-agent systems: analysis from a Lagrangian and an Eulerian perspective</i>
16h30 - 17h00	Samuel Daudin , Université Côte d'Azur <i>A Comparison Principle for Semi-Linear Hamilton-Jacobi Equations in the Wasserstein Space</i>
17h00 - 17h30	Alex Delalande , EPFL Lausanne <i>Quantitative Stability of the Pushforward Operation by an Optimal Transport Map</i>
17h30 - 18h00	Averil Prost , INSA Rouen Normandie <i>Viscosity solutions in the Wasserstein space: links between test functions and semidifferentials</i>
18h30 - 19h30	AG SMAI MODE par TERENCE Bayen
	Amphi AE1

Programme jeudi 28 mars 2024 (matin)

9h00 - 10h00	Conférence plénière
	Amphi AE1 – Session chair: Antonin Chambolle
	Jérôme Malick , CNRS <i>Nonsmooth optimization at work</i>
10h00-11h00	Session poster
Antonin Bavoil	<i>Optimisation du cycle limite d'un cerf-volant</i>
Jules Berry	<i>Approximation et perturbation de solutions stables d'un système MFG du second ordre stationnaire</i>
Mhamed Essafri	<i>Exact continuous relaxations of ℓ_0-regularized criteria with non-quadratic data terms</i>
Anne Gagneux	<i>Improving Sparse Penalties with Nonconvexity and Coefficient Clustering</i>
Joël Garde	<i>Sur l'apprentissage de combinaisons linéaires par programmation convexe sur les mesures</i>
Theo Guyard	<i>A Generic Branch-and-Bound for ℓ_0-Penalized Problems</i>
Julien Hermant	<i>Algorithmes inertiels en non convexe (quasar convexe)</i>
Thibault Moquet	<i>Sur la dualité entre les algorithmes de Frank-Wolfe et les méthodes de plans coupants</i>
Can Pouliquen	<i>Unrolled algorithms in the SPD cone and application to sparse graph learning</i>
Rémi Vaucher	<i>Estimating Signatures using ADMM for sparse third order tensor factorisation</i>
11h00 - 13h00	Inverse problems and machine learning
	Amphi AE1 – Session chair: Rémi Gribonval
11h00 - 11h30	Jonathan Chirinos Rodriguez , Università degli Studi di Genova <i>Learning the optimal regularization parameter in inverse problems</i>
11h30 - 12h00	Borjan Geshkovski , Inria, LJLL <i>Mathematical perspectives on Transformers</i>
12h00 - 12h30	Theo Guyard , INSA Rennes, INRIA Rennes <i>Screen-and-Relax for Sparse Support Identification</i>
12h30 - 13h00	Clara Lage , ENS de Lyon <i>Non-linear and non-convex sparse coding for DNA analysis in single molecule</i>
11h00 - 13h00	Games & Economy
	Salle E011 – Session chair: Bruno Ziliotto
11h00 - 11h30	Jean-Marc Bonnisseau , Centre d'économie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne <i>Incomplete markets with a countable number of states: Equilibrium and No-Arbitrage</i>
11h30 - 12h00	Maurizio D'Andrea , Luiss Guido Carli <i>Exponential Weights Algorithm with Constant Step-Size in Potential Games</i>
12h00 - 12h30	Guillaume Viger , Université Paris-Dauphine, CEREMADE <i>Zero-sum stochastic games with intermittent observation of the state</i>
11h00 - 13h00	Transport, Risk, and Control
	Salle E010 – Session chair: Nicolas Bonneel
11h00 - 11h30	Guillaume Chazareix , Inria Paris <i>Contrôle de volatilité locale par transport entropique</i>
11h30 - 12h00	Olivier Faugeras , TSE-R INRAE <i>Two Optimal Transportation approaches to measuring Risk.</i>
12h00 - 12h30	Alessandro Scagliotti , Technical University of Munich, Munich Center for Machine Learning <i>Approximations de la fonction de transport optimal par des flux de systèmes linéaires en les contrôles</i>
12h30 - 13h00	Cyprien Tamekue , Université Paris Saclay <i>Contrôlabilité et illusions visuelles</i>

Programme jeudi 28 mars 2024 (après-midi)

14h30 - 15h30 Conférence plénière

Amphi AE1 – Session chair: Simon Masnou

Yann Ollivier, Meta Fundamental AI Research, Paris

Vers un contrôleur optimal unique pour toutes les tâches: contrôle optimal universel, chaînes de Markov et deep learning

16h00 - 17h30 Session industrielle

Amphi AE1 – Session chair: Elie Bretin

16h00 - 16h30 **Sandrine Charousset** (EDF)

L'optimisation au service de la transition énergétique : quelques avancées récentes

16h30 - 17h00 **Florian Fontan** (Artelys France)

Artelys Knitro, un solveur de problèmes non-linéaires et non-linéaires en nombres entiers

17h00 - 17h30 **Eugène Ndiaye** (Apple)

From Conformal Predictions to Confidence Regions

17h45 - 18h45 Conférence grand public

Amphi Jordan – Session Chair: Filippo Santambrogio

Ivar Ekeland, CEREMADE, Université Paris-Dauphine

Quand optimiser nuit : le cas du réchauffement climatique

Programme vendredi 29 mars 2024 (matin)

9h00 - 10h00 Conférence plénière

Amphi AE1 – Session chair: T rence Bayen

Pierre Cardaliaguet, Universit  Paris Dauphine

On the optimal control of large particle systems

10h00-11h00 Session poster

- Claudia Alvarez Latuz** *Stability and optimal control of the chemostat system including a mutation term*
Adrien Cances *Transport optimal partiel et estimation de risque*
Ruben Chenevat *Optimal structures for a model of crop irrigation with constraints*
Giacomo Cozzi *Long-time asymptotics of the Sliced Wasserstein Gradient Flow*
David Lurie *Weak Ergodicity Approach to POMDPs*
Hugo Malamu *Entropic Optimal Transport Solutions of the Semigeostrophic Equations*
Lo c Marchesini *Jeux de taux de fuite et jeux de multiplication matricielle*
Camila Martinez Parra *A Two-Timescale Decision-Hazard-Decision Formulation for Storage Usage Values Calculation in Energy Systems Under Uncertainty*
Maxime Sylvestre *Existence, R gularit  et Calcul des potentiels duaux pour le transport optimal faible*
Fernanda Urrea *Analysing necessary optimality conditions for a time optimal control problem in Wasserstein spaces*
Vincent Florian *Optimal transport-based robust optimization with SkWDRO*

11h00 - 13h00 Th matiques actuelles en optimisation de forme

Amphi AE1 – Session chair: Benjamin Bogosel

- 11h00 - 11h30 **Elie Bretin**, ICJ-INSA de Lyon
Un mod le de Cahn-Hilliard d'ordre deux pour l'approximation de flots de diffusion de surface de structures fines
- 11h30 - 12h00 **Luca Gorini**,  cole Polytechnique
Optimisation g om trique de formes pour l'optique r flexive
- 12h00 - 12h30 **Aymeric Jacob de Cordemoy**, Laboratoire Jacques Louis Lions - Sorbonne Universit 
Optimisation de forme et in quations variationnelles : le probl me de Tresca dans le mod le 2D de l' lasticit  lin aire
- 12h30 - 13h00 **Idriss Mazari-Fouquer**, Paris Dauphine Universit  PSL
Is the Faber-Krahn inequality true for the Stokes operator?

11h00 - 13h00 Mean field games

Salle E010 – Session chair: Jean-Baptiste Caillaud

- 11h00 - 11h30 **Annette Dumas**, Institut Camille Jordan, Universit  Claude Bernard Lyon 1
Existence and Lipschitz regularity of the trajectories of density minimizing the total variation
- 11h30 - 12h00 **Jo o-Miguel Machado**, Universit  Paris-Dauphine
 Γ convergence of a N -player game to a Lagrangian Mean Field Game with pair-wise interaction
- 12h00 - 12h30 **Cristian Mendico**, Universit  di Roma Tor Vergata
Jeux   champ moyen avec contraintes non-holonomiques
- 12h30 - 13h00 **Ahmad Zorkot**, Universit  de Limoges
A Lagrange-Galerkin scheme for first order mean field games systems

11h00 - 13h00 Optimisation

Amphi AE2 – Session chair: Mathurin Massias

- 11h00 - 11h30 **Sadok Jerad**, Toulouse INP
M thodes d'ordre  lev es sans  valuation de la fonction objective
- 11h30 - 12h00 **Hippolyte Labarri re**, Universit  degli Studi di Genova
Inertial methods beyond minimizer uniqueness
- 12h00 - 12h30 **Guillaume Lauga**, ENS Lyon
M thodes proximales multi-niveaux pour la restauration d'images
- 12h30 - 13h00 **Iskander Sabri Legheraba**, Universit  Paris Dauphine-PSL
Optimization Landscape of Matrix Square Root Approximation

Programme vendredi 29 mars 2024 (après-midi)

14h30 - 16h30	Linear Programming (and beyond)
	Amphi AE1 – Session chair: Xavier Allamigeon
14h30 - 15h00	Sophie Huiberts , LIMOS <i>Open Problems about the Simplex Method</i>
15h00 - 15h30	Frédéric Meunier , École des Ponts <i>Don't choose between workload balance and makespan minimization</i>
15h30 - 16h00	Lionel Pournin , Université Paris 13 <i>La distance entre polyèdres en nombres entiers</i>
16h00 - 16h30	Adrian Vladu , CNRS, IRIF <i>Interior Point Methods with a Gradient Oracle</i>
14h30 - 16h30	Risk & stochastic
	Salle E010 – Session chair: Aurélien Garivier
14h30 - 15h00	Clément Royer , Université Paris Dauphine-PSL <i>Random subspaces and expected decrease in derivative-free optimization</i>
15h00 - 15h30	Atik Touazi , Université of Algiers, Benyoucef Benkhedda <i>Estimateurs à noyaux dans l'étude de la stabilité forte des modèles de risque bidimensionnels</i>
14h30 - 16h30	Set-valued maps in optimisation
	Amphi AE2 – Session chair: N. Pustelnik
14h30 - 15h00	Florent Nacry , Université Perpignan Via Domitia <i>Sous-régularité métrique et prox-régularité</i>
15h00 - 15h30	Chadi Nour , Lebanese American University <i>Numerical Method for a Controlled Sweeping Process with Nonsmooth Sweeping Set</i>
15h30 - 16h00	Sholom Schechtman , Télécom SudParis <i>The gradient's limit of a definable family of functions is a conservative set-valued field</i>
16h00 - 16h30	Michel Théra , XLIM <i>Fixed points of regular set-valued mappings in quasi-metric spaces</i>
16h30 - 17h00	Remise du prix Dodu
	Amphi AE1

On the linear convergence of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

- **Anne Auger** (INRIA, CMAP)

Mots-clé :

Résumé :

The CMA-ES algorithm is a stochastic derivative-free optimization method that can solve difficult numerical optimization problems which can typically be non-convex, ill-conditioned, or noisy. The impact of the method dramatically increased in the past few years with many applications in medicine, biology, machine learning. . .

The algorithm presents similarities to quasi-Newton techniques while the update of its state parameters does not use gradients, nor function values but only the information given by the ranking of candidate solutions on the function to optimize. In particular, the algorithm is affine-invariant and on composites of strictly increasing mappings and strictly convex-quadratic functions, in expectation, the covariance matrix of the distribution used to sample candidate solutions becomes proportional to the inverse Hessian of the strictly convex-quadratic function. The incumbent converges linearly towards the optimum.

In this presentation, I will first introduce the main ideas that lead to the CMA-ES algorithm. I will then discuss the theoretical property of affine-invariance and explain recent progress in the analyses of the linear convergence of CMA-ES. I will in particular sketch how affine-invariance and the stability analysis of an underlying Markov chain implies the linear convergence of the algorithm and the learning of the inverse Hessian (on strictly convex-quadratic functions).

This work is in collaboration with Armand Gissler (PhD student, CMAP, Ecole Polytechnique) and Nikolaus Hansen (Inria and CMAP, Ecole Polytechnique).

Deep Unfolding Approach for Limited-Angle Computed Tomography Image Reconstruction

- **Emilie Chouzenoux** (INRIA, CVN)

Mots-clé :

Résumé :

This talk presents recent developments made in the context of a PhD CIFRE thesis between CVN lab, in CentraleSupélec, Univ. Paris Saclay, and GE Healthcare R&D department. We addressed the inverse problem arising in computational imaging, of the regions of interest (ROI) reconstruction from a limited number of computed tomography (CT) measurements. Classical model-based iterative reconstruction methods lead to images with predictable features. Still, they often suffer from tedious parameterization and slow convergence. On the contrary, deep learning methods are fast, and they can reach high reconstruction quality by leveraging information from large datasets, but they lack interpretability. At the crossroads of both methods, deep unfolding networks have been recently proposed. Their design includes the physics of the imaging system and the steps of an iterative optimization algorithm. Motivated by the success of these networks for various applications, we introduced in [1] an unfolding neural network designed for ROI CT reconstruction from limited data. Few-view truncated data are effectively handled thanks to a robust non-convex data fidelity term combined with a sparsity-inducing regularization function. We unfold a block alternating proximal algorithm, embedded in an iterative reweighted scheme, allowing the learning of key parameters in a supervised manner. Our experiments showcase an improvement over several state-of-the-art methods, including a model-based iterative scheme, a multi-scale deep learning architecture, and other deep unfolding networks.

Références

- [1] M. Savanier, E. Chouzenoux, J.-C. Pesquet, C. Riddell Deep Unfolding of the DBFB Algorithm with Application to ROI CT Imaging with Limited Angular Density. *IEEE Transactions on Computational Imaging*, vol. 9, pp. 502-516, 2023

Nonsmooth optimization at work

- Jérôme Malick (CNRS, LJK)

Mots-clé :

Résumé : Nonsmoothness pervades the realm of optimization, from e.g. relaxations of combinatorial optimization problems to robust models, and deep learning activation functions. In this broad spectrum of applications, nonsmoothness isn't random; instead, nonsmooth objective functions often exhibit a significant structure, driving algorithms and bestowing fundamental properties upon their minimizers. In this talk, I will provide a comprehensive overview of nonsmoothness in optimization, emphasizing its nice geometric properties, and demonstrating its impact on practical applications, particularly in machine learning and energy management. Additionally, I will put into perspective recent advancements within the historical framework of convex and variational analysis, pioneered by J.-J. Moreau.

Vers un contrôleur optimal unique pour toutes les tâches: contrôle optimal universel, chaînes de Markov et deep learning

- Yann Ollivier (Meta AI, Paris)

Mots-clé :

Résumé :

Le contrôle optimal et l'apprentissage par renforcement cherchent à produire des règles de comportement qui optimisent un certain objectif prédéfini dans un environnement (par exemple, un robot qui doit maximiser sa vitesse). Nous considérons ici un problème plus ambitieux : peut-on produire un agent qui saura optimiser n'importe quel nouvel objectif en temps réel, même s'il n'a pas été construit spécifiquement pour cet objectif ?

Plus formellement : peut-on manipuler en pratique la fonction qui, à un objectif de contrôle optimal, associe le comportement maximisant cet objectif ? Cela permettrait de s'adapter en temps réel à de nouveaux objectifs. Difficulté supplémentaire : les lois physiques de l'environnement ne sont pas connues, et on doit juste travailler à partir d'exemples de trajectoires observées.

Notre démo <https://controllable-agent.metademolab.com/> donne l'exemple d'un robot virtuel capable produire instantanément un comportement approximativement optimal quand on lui spécifie une fonction à optimiser, telle que "moment angulaire * vitesse horizontale". Nous expliquerons les mathématiques derrière cette construction, qui mêlent contrôle optimal, spectre des chaînes de Markov, apprentissage par renforcement et réseaux de neurones.

Quand optimiser nuit: le cas du réchauffement climatique

- **Ivar Ekeland** (Université Paris-Dauphine, Ceremade)

Mots-clé :

Résumé : En 2018, Nordhaus obtenait le prix Nobel d'économie pour avoir démontré qu'un réchauffement de 3,5°C en 2100 était optimal. Le consensus scientifique (économistes mis à part) est qu'au delà de 2°C de réchauffement, on va vers une catastrophe. La démonstration de Nordhaus est parfaitement rigoureuse, et repose sur des bases bien établies. Je m'interrogerai sur la pertinence du concept d'optimisation pour modéliser le comportement humain, et je proposerai d'autres pistes

On the optimal control of large particle systems

- **Pierre Cardaliaguet** (Université Paris-Dauphine, Ceremade)

Mots-clé :

Résumé : The optimal control of large particle systems has known a growing interest in the recent years, in the line of mean field games. The goal of this talk is to present some recent advances on the subject (large number limit, convergence rate, link with Hamilton-Jacobi equations in the space of measure), together with some applications.

Extending convexity and gradient descent: a framework for general costs based on alternating minimization

- Flavien Léger (INRIA Paris, MOKAPLAN)
- Pierre-Cyril Aubin-Frankowski (TU Wien, VADOR)

Mots-clé : continuous optimization, gradient descent, general cost

Résumé : In this talk we go beyond the quadratic cost in optimization, by replacing it with a general cost function and using a majorize-minimize framework. We unveiled in [1] a new class of gradient-type optimization methods that extends vanilla gradient descent, mirror descent, Riemannian gradient descent, and natural gradient descent, while keeping the same proof ideas and (sub)linear rates of convergence. Our approach involves constructing a surrogate for the objective function f in a systematic manner, based on a chosen cost function c . This surrogate $\phi(x, y) = c(x, y) + f^c(y)$ is then minimized using an alternating minimization scheme. We obtain novel rates of convergence for this general algorithm based on Csiszar and Tusnady’s “five-points inequality” [2].

Specifying the results to cost functions c , using optimal transport theory we establish convergence rates based on generalized notions of L -smoothness and convexity, namely c -concavity and the c -cross-convexity that we introduce. We provide local versions of these two notions when the cost satisfies a condition known as nonnegative cross-curvature. In particular our framework encompasses the concept of relative smoothness introduced in 2017 in [3] for Bregman divergences, and we provide the first global rates for natural gradient descent and Newton’s method under a new assumption distinct from self-concordance.

Références

- [1] F. Léger and P.-C. Aubin-Frankowski. Gradient descent with a general cost. arxiv.org/abs/2305.04917, 2023.
- [2] I. Csiszár and G. Tusnády. Information Geometry and Alternating Minimization Procedures. *Statistics and Decisions*, 205–237, 1984.
- [3] H. H. Bauschke, J. Bolte and M. Teboulle. A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017.

Screen-and-Relax for Sparse Support Identification

- Théo Guyard (Inria and Insa, CNRS, IRMAR - UMR CNRS 6625, Rennes, France)
- Cédric Herzet (Ensaï, CNRS, CREST - UMR 9194, Rennes, France)
- Clément Elvira (CentraleSupélec, CNRS, IETR - UMR 6164, Rennes, France)

Mots-clé : Sparse problems, screening tests, relaxing tests.

Résumé : Sparse modeling arises in various application fields such as machine learning, signal processing and inverse problems, among others [1]. In this context, it is of tantamount interest to solve convex optimization problems of the form

$$\mathbf{x}^* \in \arg \min_{\mathbf{x}} f(\mathbf{x}) + \lambda g(\mathbf{x})$$

where $f(\cdot)$ is a loss function and $g(\cdot)$ is a sparsity-inducing regularization whose strength is driven by the parameter $\lambda > 0$. The latter problem can be addressed with first-order optimization methods such as Proximal Gradient (PG), Coordinate Descent (CD) or Alternating Direction Method of Multipliers (ADMM) [2]. A popular strategy to exploit the problem structure is to enhance these algorithms with *screening tests* to detect the position of *zeros* in \mathbf{x}^* . This allows shrinking the problem dimensionality and saving computational resources during the resolution process [3]. We propose a complementary strategy, dubbed *relaxing*, to detect the position of *non-zeros* in \mathbf{x}^* . Leveraging this knowledge, we partially smooth the objective function of the problem which allows performing second-order iterations instead of first-order ones. In particular, this allows reaching *super-linear* convergence rates in finite time although only *sub-linear* or *linear* convergence rates are typically accessible for such optimization problems. By implementing screening and relaxing tests together in any solution method, we thus expect a double acceleration resulting from the dimensionality reduction and the convergence rate improvement. The proposed method is assessed on both synthetic and real-world datasets of varying characteristics and we observe an acceleration factor of up to three orders of magnitude compared to a vanilla implementation of the PG, CD and ADMM algorithms.

Références

- [1] Tropp, J. et al., Computational methods for sparse solution of linear inverse problems, *Proceedings of the IEEE*, 98(6), 948-958, 2010.
- [2] Bach, F. et al., Convex optimization with sparsity-inducing norms *Optimization for machine learning*, 5, 19-53, 2011.
- [3] Xiang, Z. et al., Screening tests for lasso problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5), 1008-1027, 2016.

A Generic Branch-and-Bound for ℓ_0 -Penalized Problems

- **Théo Guyard** (Inria and Insa, CNRS, IRMAR - UMR CNRS 6625, Rennes, France)
- Cédric Herzet (Ensaï, CNRS, CREST - UMR 9194, Rennes, France)
- Clément Elvira (CentraleSupélec, CNRS, IETR - UMR 6164, Rennes, France)
- Ayşe-Nur Arslan (Inria, CNRS, IMB - UMR 5251 Bordeaux, France, France)

Mots-clé : Sparse modeling, ℓ_0 -penalized problems, branch-and-bound.

Résumé : Sparse modeling arises in various applied mathematical fields such as machine learning or signal processing [1]. In this context, addressing ℓ_0 -penalized problems of the form

$$\mathbf{x}^* \in \arg \min_{\mathbf{x}} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

is of particular interest. The function $f(\cdot)$ is a loss linked to the considered model, the ℓ_0 -norm enforces sparsity in the optimizer by counting the number of non-zeros in its input and the penalty term $h(\cdot)$ promotes other desirable properties that may be relevant for the application at hand. Problems involving the ℓ_0 -norm are NP-hard in the general case but they can be tackled using mixed-integer optimization methods. On the one hand, off-the-shelf solvers can handle a large variety of instances [2] but suffer from bad numerical performance since they are not fully able to exploit the sparse structure of the problem. On the other hand, specialized Branch-and-Bound (BnB) methods achieve more reasonable solving time but can only address specific instances of the problem, namely with a quadratic loss function and a penalty term either encoding a Big-M constraint [3] and/or an ℓ_2 -norm [4].

We propose a generic framework for specialized BnB solvers that can handle any instance of ℓ_0 -penalized problems verifying some mild hypotheses and which achieves competitive numerical performance. Through new theoretical results, we show that all the ingredients required for its implementation can be obtained in closed form, even though we do not require the explicit expression of $f(\cdot)$ and $h(\cdot)$. Interestingly, we show that our approach encompasses prior works as a special case. Capitalizing on the genericity of our framework, we provide **E10ps**, a flexible implementation of our BnB solver with a plug-and-play workflow allowing user-defined functions $f(\cdot)$ and $h(\cdot)$ to best fit the practitioner's needs. Through different experiments, we show that our solver achieves state-of-the-art performance on classical instances of the problem and can address new instances for which no specialized solvers exist so far and that have therefore been out of computational reach. Our contribution thus paves the way for new application opportunities.

Références

- [1] Tropp, J. et al., Computational methods for sparse solution of linear inverse problems, *Proceedings of the IEEE*, 98(6), 948-958, 2010.
- [2] Bourguignon, S. et al., Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance, *IEEE Transactions on Signal Processing*, 64(6), 1405-1419, 2016.
- [3] Ben Mhenni, R. et al., Global optimization for sparse solution of least squares problems, *Optimization Methods and Software*, 37(5), 1740-1769, 2022.
- [4] Hazimeh, H., et al., Sparse regression at scale: Branch-and-bound rooted in first-order optimization, *Mathematical Programming*, 196(1-2), 347-388, 2022.

Weak Ergodicity Approach to POMDPs

- Krishnendu Chatterjee (Institute of Science and Technology Austria)
- **David Lurie** (CEREMADE, Université Paris Dauphine, PSL Research Institute, NyxAir)
- Raimundo Saona (Institute of Science and Technology Austria)
- Bruno Ziliotto (CEREMADE, CNRS, Université Paris Dauphine, PSL Research Institute)

Mots-clé : Markov, approximation algorithms, weak ergodicity

Résumé : We consider discrete-time partially observable Markov decision processes (POMDPs) with the long-run average objective. This model serves as a powerful tool for sequential decision-making problems, effectively accounting for both the dynamic of state transitions and the partial information of the system. In this work, we introduce new classes of POMDPs leveraging the concept of “weak ergodicity” and prove their decidability.

Addressing POMDPs with the long-run average objective presents significant challenges. Computationally, finite horizon POMDPs are identified as PSPACE-complete [1], while infinite horizon POMDPs are known as undecidable [2]. The analytical difficulties of the long-run average objective is mainly due to the absence of the contraction property. Consequently, this work aims to develop innovative algorithms to overcome these difficulties

We propose a new approach which relies on the property of “weak ergodicity” in products of matrices. This concept [3] describes how a system progressively forgets its past. These techniques have found wide applications in the literature, including demographic studies and economic analysis [4]. As a result, the insights from our study suggest that weak ergodicity theory might serve as a promising direction for implementing new efficient algorithms to address the theoretical challenges associated with POMDPs.

Références

- [1] Christos H Papadimitriou and John N Tsitsiklis. The complexity of Markov decision processes. *Mathematics of operations research*, 12:441–450, 1987.
- [2] Omid Madani and Steve Hanks and Anne Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147:5–34, 2003.
- [3] Andrei Kolmogoroff. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–458, 1931.
- [4] Eugene Seneta. Non-negative matrices and Markov chains. *Springer Science & Business Media*, 2006.

Saddle-to-Saddle Dynamics in Diagonal Linear Networks

- Scott Pesme (EPFL)
- Nicolas Flammarion (EPFL)

Mots-clé : Saddle-to-Saddle Dynamics, Incremental Learning, Diagonal Linear Networks

Résumé : When training neural networks with gradient methods using a small initialisation of the weights, strange types of learning curves appear: the training process makes very little progress for some time, followed by a sharp transition where a new “feature” is suddenly learnt (Figure 1). This behaviour is usually referred to as incremental learning. In this talk, I will show that we can fully describe this phenomenon when considering a toy network architecture which corresponds to a diagonal linear network. In this simplified setting, we can prove that the gradient flow trajectory jumps from a saddle of the training loss to another. Each visited saddle as well as the jump times can be computed through a recursive algorithm reminiscent of the Homotopy algorithm used for finding the Lasso path. The talk is based on our paper [1].

Saddle-to-saddle dynamics:

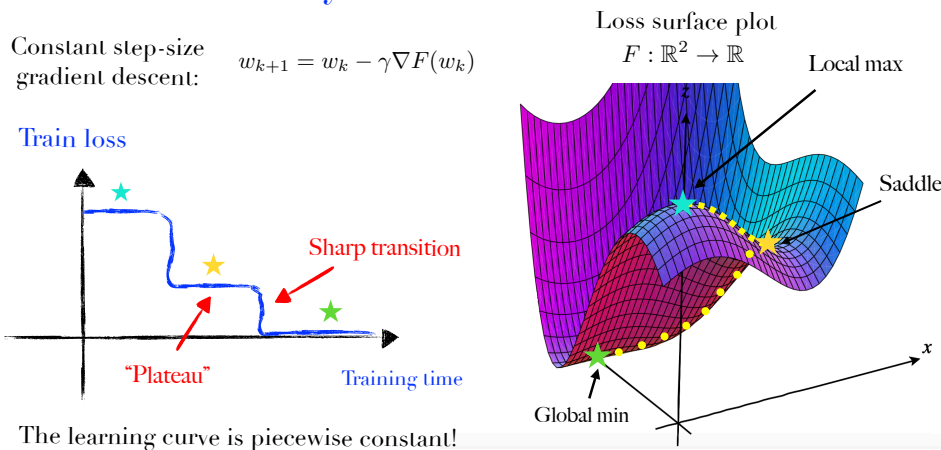


Figure 1: Toy illustration of saddle-to-saddle dynamics: the optimisation trajectory jumps from a critical point of the training loss to another, until reaching a global minimum.

Références

- [1] Scott Pesme and Nicolas Flammarion. Saddle-to-Saddle Dynamics in Diagonal Linear Networks. *Advances in Neural Information Processing Systems*, 2023.

Jeux á champ moyen avec contraintes non-holonomiques

- Cannarsa Piermarco (Università di Roma Tor Vergata)
- Mannucci Paola (Università degli studi di Padova)
- Marchi Claudio (Università degli studi di Padova)
- **Mendico Cristian** (Università di Roma Tor Vergata)

Mots-clé : Jeux á champ moyen, Contraintes non-holonomiques, Équations subelliptiques

Résumé : Dans cet exposé, nous concentrons notre attention sur le système de jeux á champ moyen déterministe et stochastique avec des contraintes non-holonomiques. De telles contraintes sont modélisées conférant á l'espace euclidien une structure sous-riemannienne. En particulier, au cours de l'exposé, nous nous concentrons sur les groupes de Lie homogènes.

Dans le cas stochastique, nous présenterons d'abord des résultats bien posés pour les équations subelliptiques de Hamilton-Jacobi et pour l'équation de Fokker-Planck, fournissant également une interprétation probabiliste. Ensuite, nous appliquons de tels résultats prouvant l'existence de solutions classiques—au sens horizontal—pour un horizon de temps court.

Considérant le cas du premier ordre, nous présentons d'abord quelques résultats récents sur l'approche d'Aubry-Mather du comportement asymptotique des solutions de l'équation de Hamilton-Jacobi sur des groupes de Lie homogènes. Ensuite, nous appliquons ces résultats pour étudier le comportement asymptotique des solutions de un système de jeux á champ moyen. En tant que sous-produit du comportement asymptotique, nous présentons quelques résultats sur la contrôlabilité de l'équation de transport.

Références

- [1] Paola Mannucci, Claudio Marchi, Cristian Mendico, Semi-linear parabolic equations on homogeneous Lie groups arising from mean field games. *arXiv:2307.07257*
- [2] Piermarco Cannarsa, Cristian Mendico, Weak KAM Theory and Aubry-Mather Theory for sub-Riemannian control systems. *arXiv:2306.03808*
- [3] Lucia Caramellino, Cristian Mendico, Subelliptic Fokker-Planck equations and probabilistic counterparts. *Work in progress.*

Jeux de taux de fuite et jeux de multiplication matricielle

- Marianne Akian (Inria and CMAP, École polytechnique, CNRS, IP Paris)
- Stéphane Gaubert (Inria and CMAP, École polytechnique, CNRS, IP Paris)
- **Loïc Marchesini** (CMAP, École polytechnique, CNRS, IP Paris and Inria)

Mots-clé : Jeux à Paiement Moyen, Fonctions non-expansives, Rayon Spectral Conjoint

Résumé : Nous introduisons une nouvelle classe de jeux répétés à somme nulle où les joueurs cherchent à minimiser ou à maximiser le taux de fuite d'un système dynamique partiellement contrôlé, dont l'évolution temporelle est déterminée par des opérateurs non-expansifs. Le taux de fuite est mesuré à l'aide d'une *hemi-distance*, c'est-à-dire d'une fonction de deux variables vérifiant l'inégalité triangulaire, mais non nécessairement symétrique ou positive.

Formellement, on se donne deux ensembles d'actions compacts \mathcal{A} et \mathcal{B} , un espace d'états X muni d'une hémis-distance d , un état initial $x_0 \in X$ et une famille d'opérateurs $(T_{ab})_{(a,b) \in \mathcal{A} \times \mathcal{B}}$ non-expansifs relativement à d , c'est-à-dire tels que $d(T_{ab}(x), T_{ab}(y)) \leq d(x, y)$ pour tous $x, y \in X$. Deux joueurs, appelés *Min* et *Max*, choisissent à tour de rôle une action, respectivement dans les ensembles \mathcal{A} et \mathcal{B} . Le jeu se joue en information parfaite. Étant donné une suite d'actions $(a_n, b_n)_{n \in \mathbb{N}^*}$, le *taux de fuite* est défini comme suit

$$\limsup_{k \rightarrow \infty} \frac{d(T_{a_k b_k} \circ \cdots \circ T_{a_1 b_1}(x_0), x_0)}{k} .$$

Le joueur *Min* cherche à minimiser ce critère et le joueur *Max* cherche à le maximiser. En étudiant l'opérateur

$$S\nu(x) := \inf_{a \in \mathcal{A}} \sup_{b \in \mathcal{B}} \nu(T_{ab}(x))$$

agissant sur l'espace des fonctions 1-Lipschitziennes de (X, d) dans (\mathbb{R}, h) où h est l'hémis-distance $(x, y) \mapsto x - y$, nous démontrons l'existence de la *valeur* pour notre jeu, que celle-ci est égale à la limite de la valeur du jeu en horizon fini, et qu'elle est en outre une *valeur uniforme* [1]. De plus, nous établissons une caractérisation de type *Collatz-Wielandt* de cette valeur:

$$\text{val} = \max\{\lambda \in \mathbb{R} \mid \exists \nu \text{ 1-Lipschitzienne, } \lambda + \nu \leq S\nu\}$$

Enfin, nous démontrons l'existence de stratégies optimales pour les deux joueurs.

En considérant l'hémis-distance sur l'intérieur du cône positif de \mathbb{R}^n , $d(x, y) := \log \max_i \frac{x_i}{y_i}$, et en prenant des opérateurs de la forme $T_{ab}(x) = xM_{ab}$, où M_{ab} est une matrice, nous déduisons comme corollaire l'existence d'une valeur uniforme pour les *jeux de multiplication matricielle* introduits par Asarin et al. [2], lorsque les matrices sont à coefficients positifs. Dans le cas où il y a un seul joueur, et où ce joueur est le maximiseur, la valeur du jeu coïncide avec le logarithme du rayon spectral conjoint ([3]).

Références

- [1] Mertens, J.-F., A. Neyman. 1981. Stochastic games. *Internat. J. Game Theory* **10**(2) 53–66.
- [2] E. Asarin, J. Cervelle, A. Degorre, C. Dima, F. Horn, V. Kozyakin. Entropy Games and Matrix Multiplication Games. Proc. of STACS'2016.
- [3] Gian-Carlo Rota, W. Gilbert Strang. A note on the joint spectral radius. *Indagationes Mathematicae*, 1960.

Learning the optimal regularization parameter in inverse problems

- **Jonathan Chirinos Rodriguez** (MaLGa, DIMA, University of Genova)
- Ernesto De Vito (MaLGa, DIMA, University of Genova)
- Cesare Molinari (MaLGa, DIMA, University of Genova)
- Lorenzo Rosasco (MaLGa, DIBRIS, University of Genova)
- Silvia Villa (MaLGa, DIMA, University of Genova)

Mots-clé : regularization of inverse problems, supervised learning, data-driven approaches

Résumé : Selecting the best regularization parameter in inverse problems is a classical and yet challenging problem. Recently, data-driven approaches [1] have become popular to tackle this challenge due to its promising results in many applications [3, 4]. Nevertheless, few theoretical guarantees have been provided up to date.

In this work, we provide a theoretical analysis for this problem by applying classical statistical learning techniques [2]. In particular, we characterize the error performance of this method following an empirical risk minimization approach. We show that, provided with enough data, this approach can reach sharp rates while being essentially adaptive to the noise and smoothness of the problem. Our analysis studies a wide variety of regularization methods, including spectral regularization methods (Tikhonov regularization, Landweber iteration, the ν -method), non-linear Tikhonov regularization and general convex regularizers such as sparsity inducing norms and Total Variation regularization. Finally, we show some numerical simulations that corroborate and illustrate the theoretical findings.

Références

- [1] Arridge, S., Maass, P., Öktem, O. and Schönlieb, C.-B.. Solving inverse problems using data-driven models. *Acta Numerica*, 1–174, 2019.
- [2] Cucker, F. and Smale, S.. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39:1–49, 2002.
- [3] Kobler, E., Effland, A., Kunisch, K. and Pock, T.. Total Deep Variation for Linear Inverse Problems. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7546–7555, 2020.
- [4] Kunish, K. and Pock, T.. A Bilevel Optimization Approach for Parameter Learning in Variational Models. *SIAM Journal on Imaging Sciences*, 6:938–983, 2013.

A Two-Timescale Decision-Hazard-Decision Formulation for Storage Usage Values Calculation in Energy Systems Under Uncertainty

- Camila Martinez Parra (RTE-France)
- Jean-Marc Janin (RTE-France)
- Manuel Ruiz (RTE-France)
- Michel De Lara (Cermics, École des Ponts, France)
- Jean-Philippe Chancelier (Cermics, École des Ponts, France)
- Pierre Carpentier (UMA, ENSTA Paris)

Mots-clé : Energy system modelling, information structure, stochastic multistage optimization, dynamic programming.

Résumé : The penetration of renewable energies requires additional storages to deal with the system’s intermittency. Accordingly, there is growing interest in evaluating the opportunity cost associated with stored energy (usage value) which can be obtained by solving a multi-stage stochastic optimization problem. Today, to compute usage values under uncertainties, an adequacy resource problem is solved using stochastic dynamic programming assuming a *hazard-decision* information structure. This modelling assumes completely knowledge of the coming week uncertainties; this is not adapted to the system operation, as the intermittency occurs at smaller timescale. We propose a two-timescale problem formulation with a new information structure considering planning and recourse decisions: *decision-hazard-decision*. This structure is used to decompose the multistage decision-making process into a nonanticipative planning step (in which the switch-on decisions for the thermal units are made), and a recourse step (in which the power modulation decisions are made once the uncertainties have been disclosed). We discuss numerical resolution.

Références

- [1] R. E. Bellman, *Dynamic Programming*. Princeton, N.J.: Princeton University Press, 1957.
- [2] P. Carpentier, J.-P. Chancelier, M. D. Lara, T. Martin, and T. Rigaut, “Time blocks decomposition of multistage stochastic optimization problems,” *Journal of Convex Analysis*, in press.
- [3] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.
- [4] G. Steeger, L. Barroso, and S. Rebennack, “Optimal bidding strategies for hydro-electric producers: A literature survey,” *Power Systems, IEEE Transactions on*, vol. 29, pp. 1758–1766, 07 2014.

(S)GD over Diagonal Linear Networks: Implicit Regularisation, Large Stepsizes and Edge of Stability

- **Mathieu Even** (Inria - ENS Paris)

Mots-clé :

Résumé : In this work, we investigate the impact of stochasticity and large stepsizes on the implicit regularisation of gradient descent (GD) and stochastic gradient descent (SGD) over diagonal linear networks. We prove the convergence of GD and SGD with macroscopic stepsizes in an overparametrised regression setting and characterise their solutions through an implicit regularisation problem. Our crisp characterisation leads to qualitative insights about the impact of stochasticity and stepsizes on the recovered solution. Specifically, we show that large stepsizes consistently benefit SGD for sparse regression problems, while they can hinder the recovery of sparse solutions for GD. These effects are magnified for stepsizes in a tight window just below the divergence threshold, in the “edge of stability” regime. Our findings are supported by experimental results.

Potentiels convexes incertains et aversion au risque pour la mécanique des matériaux

- **Jérémy Bleyer** (Laboratoire Navier, Ecole des Ponts ParisTech, Univ Gustave Eiffel, CNRS)

Mots-clé : programmation stochastique, potentiels convexes, mesure de risque, mécanique des matériaux, approche variationnelle

Résumé : Ce travail s'intéresse à la formulation d'un comportement effectif pour les lois de comportement des matériaux en présence d'incertitudes. Nous introduisons une formulation variationnelle thermodynamique du comportement incertain dans le cadre de la programmation stochastique. Le comportement effectif est obtenu grâce à l'utilisation d'une mesure de risque cohérente, préservant ainsi les propriétés de convexité initiales des potentiels thermodynamiques et du cadre variationnel associé.

L'utilisation de la *Conditional Value-at-Risk* (CVaR) fournit une estimation optimiste du comportement du matériau, le rendant plus rigide et plus résistant que le cas nominal. Cependant, du point de vue des applications en ingénierie, il est souvent plus intéressant de disposer d'une estimation pessimiste, indiquant un matériau plus souple et moins résistant que le cas nominal, dans un souci de sécurité accru.

Pour répondre à cette nécessité, nous proposons une nouvelle mesure de risque cohérente, duale de la CVaR. Cette approche repose sur l'utilisation de la transformée polaire pour les fonctions convexes à valeurs positives.

Références

- [1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.
- [2] J. Bleyer. Risk-averse estimates of effective properties in heterogeneous elasticity. *Comptes Rendus. Mécanique*, 351(G1):29–42, 2023.
- [3] A. Mielke. Evolution of rate-independent systems. *Evolutionary equations*, 2:461–559, 2005.
- [4] V. Milman and S. Artstein-Avidan. Hidden structures in the class of convex functions and a new duality transform. *Journal of the European Mathematical Society*, 13(4):975–1004, 2011.
- [5] R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Contrôlabilité et illusions visuelles

- **Cyprien Tamekue** (Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France) et Institut Polytechnique des Sciences Avancées (IPSA), 63 boulevard de Brandebourg, 94200 Ivry-sur-Seine, France
- Dario Prandi (Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France)
- Yacine Chitour (Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France)

Mots-clé : Contrôle en neurosciences, Modèle de champ neuronal, Illusions visuelles et perception, Effet MacKay, Formation forcée de motifs spatiotemporels.

Résumé : Cet exposé porte sur l'étude de la relation complexe entre la perception visuelle et la modélisation mathématique de l'activité neuronale dans le cortex visuel primaire (V1). Il se concentrera principalement sur la réplification de l'effet MacKay [Mackay, Nature 1957] et des expériences psychophysiques de Billock et Tsou [PNAS, 2007]. Bien que la théorie des bifurcations ait été une approche mathématique dominante pour aborder les problèmes en neurosciences, en particulier pour décrire les formations spontanées de motifs dans V1 dues à des changements de paramètres, elle rencontre des défis dans des scénarios avec des entrées sensorielles localisées. Cela est évident, par exemple, dans les expériences psychophysiques de Mackay et de Billock et Tsou, où la redondance des informations des stimuli visuels ou le fait que ceux-ci soient localisés dans le champ visuel entraîne des formes irrégulières, rendant la théorie des bifurcations et l'analyse multi-échelle moins efficaces. Pour aborder cela, nous adoptons un point de vue mathématique basé sur la contrôlabilité entrée-sortie d'un modèle de champs neuronaux de type Amari. Dans cette approche, on considère l'entrée sensorielle comme une fonction de contrôle, la représentation corticale via la carte rétino-corticale du stimulus visuel qui capture les caractéristiques distinctes du stimulus, en particulier la redondance centrale dans le motif de l'entonnoir de MacKay "rayons de MacKay". Du point de vue de la théorie du contrôle, la propriété de contrôlabilité exacte de l'équation de type Amari est discutée pour les fonctions de réponse linéaires et non linéaires. Puis, appliquée à la réplification de l'effet MacKay, nous ajustons le paramètre représentant la connectivité intra-neuronale pour assurer que, en l'absence d'entrée sensorielle, l'activité corticale se stabilise exponentiellement à l'état stationnaire que nous étudions de manière quantitative et qualitative pour montrer qu'elle capture toutes les caractéristiques essentielles de l'image rémanente induite rapportée par MacKay et par Billock et Tsou.

Références

- [1] C. Tamekue, D. Prandi, Y. Chitour. On the mathematical replication of the MacKay effect from redundant stimulation. *arXiv: 2311.07338 [math.OA]*, 2023.
- [2] C. Tamekue, D. Prandi, Y. Chitour. Cortical origins of MacKay-type visual illusions: A case for the non-linearity. *IFAC-PapersOnLine. 22nd IFAC World Congress*, 56.2:476–481, 2023.

Existence and Lipschitz regularity of the trajectories of density minimizing the total variation

- **Annette Dumas** (ICJ - Lyon 1)
- **Filippo Santambrogio** (ICJ - Lyon 1)

Mots-clé : Mean Field Games, Jumps

Résumé : The problem I will present is motivated by the study of a Mean Field Games model whose theory was simultaneously introduced by Lasry and Lions ([1, 2]) and Caines, Huang and Malhamé ([3]) in 2006. The model consists in studying a population in a city where each agent jumps to move from one place to another. Each inhabitant minimizes a cost composed of the number of jumps and an increasing function of the density of the population. The solution to this problem is a probability measure on the trajectories which is a Nash equilibrium.

The probability measure Q on the trajectories can be seen as a trajectory of the density of the population $\rho_t = e_t \# Q$ which leads us to the following variational problem:

$$\min_{\substack{\rho \in E, \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \int_0^T \int_{\Omega} (|\partial_t \rho(t, x)| + V(t, x) \rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)),$$

where $E = BV([0, T], L^1(\Omega)) \cap L^2([0, T] \times \Omega)$. The number of jumps is associated with the total variation of the density and the additional cost is associated with the increasing function of the density. Density constraints are also added to the problem. We will see that the solution exists, is unique and is Lipschitz in time, despite the discontinuous trajectories taken by each agent. With additional hypothesis on the data, boundedness or continuity in space can be obtained with Dirichlet conditions in time.

The aspect of the solutions are given by the Euler-Lagrange equations which show that in space, either the solution is constant, or it follows the critical points of the cost. Numerical simulations are carried out on a simple example by using the Fast Dual Proximal Gradient method which validates the theoretical framework.

Références

- [1] J.-M. Lasry, P.-L. Lions. Jeux à champ moyen. I - Le cas stationnaire. *C. R. Acad. Sci. Paris, Ser. I* 343, 2006.
- [2] J.-M. Lasry, P.-L. Lions. Jeux à champ moyen. II - Horizon fini et contrôle optimal. *C. R. Acad. Sci. Paris, Ser. I* 343, 2006.
- [3] M. Huang, R. P. Malhamé and P. E. Caines Large Population Stochastic Dynamic Games: Closed-Loop McKean-Vlasov Systems and the Nash Certainty Equivalence Principle. *Communications in Information and Systems, Vol. 6, No. 3*, pp. 221-252, 2006.

Artelys Knitro, un solveur de problèmes non-linéaires et non-linéaires en nombres entiers

- **Florian Fontan** (Artelys France)

Mots-clé : programmation non-linéaire, programmation non-linéaire en nombres entiers

Résumé : Artelys Knitro est un solveur de problèmes non-linéaires et non-linéaires en nombres entiers. Il supporte les structures linéaires, quadratiques et les fonctions boîte noire, données si possible, avec leur dérivées première et seconde. Pour résoudre les problèmes non-linéaires continus, Artelys Knitro inclut deux algorithmes de point intérieur et deux algorithmes d'active set. Ces algorithmes garantissent l'optimalité globale de la solution renvoyée dans le cas d'un problème convexe, et l'optimalité locale dans le cas d'un problème non-convexe. Pour résoudre les problèmes non-linéaires en nombres entiers, Artelys Knitro inclut un algorithme de branch-and-bound non-linéaire – qui garantit l'optimalité locale de la solution pour un problème convexe – et un algorithme mixed-integer sequential quadratic programming utile pour des problèmes avec des évaluations coûteuses.

Dans cette présentation, nous commencerons par une rapide introduction à la programmation non-linéaire et aux différentes techniques développées pour résoudre les différents types de problèmes qui en sont issus. Nous verrons plus en détails les algorithmes implémentés dans Knitro et leur utilisation à travers plusieurs exemples de cas d'applications. Nous présenterons également les derniers développements qui y ont été intégrés.

Viscosity solutions in the Wasserstein space: links between test functions and semidifferentials

- **Averil Prost** (INSA Rouen Normandie)
- Hasnaa Zidani (INSA Rouen Normandie)

Mots-clé : Wasserstein space, viscosity solutions, Mayer problem

Résumé : We consider a Mayer problem posed in the space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures with finite second moment, endowed with the Wasserstein distance. Our aim is to write an Hamilton-Jacobi-Bellman equation whose unique solution is the value function of the control problem. This type of problem usually admits nonsmooth solutions, and is treated with the theory of viscosity solutions [1].

To extend the theory of viscosity solutions to $\mathcal{P}_2(\mathbb{R}^d)$, several lines of investigation are open. One can use semidifferentials [2, 3] or implement metric viscosity techniques in the case of Eikonal-type equations [5, 4]. In this talk, we follow an idea raised in [6] and consider instead the map of directional derivatives as our elementary local approximation of a smooth function, instead of a gradient. This framework allow to define viscosity solutions using quite natural test functions, and supports a strong comparison principle while retaining the link between Hamilton-Jacobi-Bellman equations and control problems [7]. We are interested in understanding in which case can we extend the natural equivalence between semidifferentials and test function definition, that stands in the finite-dimensional setting.

Références

- [1] M. G.Crandall, H. Ishii and P-L. Lions. User's Guide to Viscosity Solutions of Second Order Partial Differential Equations. *Bulletin of the American Mathematical Society*, 1992.
- [2] A. Marigonda and M. Quincampoix. Mayer control problem with probabilistic uncertainty on initial positions. *Journal of Differential Equations* 264, 2018.
- [3] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 2019.
- [4] Y. Giga, N. Hamamuki, and A. Nakayasu. Eikonal equations in metric spaces. *Transactions of the American Mathematical Society*, 2015.
- [5] L. Ambrosio and J. Feng. On a class of first order Hamilton-Jacobi equations in metric spaces. *Journal of Differential Equations*, 2014.
- [6] F. Jean, O. Jerhaoui, and H. Zidani. Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition (in press). *SIAM J. Mathematical Analysis* (in press), 2023.
- [7] O. Jerhaoui, A. Prost and H. Zidani. Viscosity solutions of centralized control problems in measure spaces (submitted). 2023.

Inertial methods beyond minimizer uniqueness

- Jean-François Aujol (Institut de Mathématiques de Bordeaux)
- Charles Dossal (INSA Toulouse)
- **Hippolyte Labarrière** (Università degli Studi di Genova)
- Aude Rondepierre (INSA Toulouse)

Mots-clé : Optimisation convexe, méthodes inertielles, conditions de croissance, unicité du minimiseur

Résumé : When considering a convex composite minimization problem, it is well known that momentum allows first-order methods to be accelerated both theoretically and numerically. Since the seminal works of Polyak in 1964 [1] and Nesterov in 1983 [2], inertial methods have been widely studied and several approaches have been proposed to study them, including the continuous analysis proposed by Su, Boyd and Candès [3] to study Nesterov’s accelerated gradient method or the Performance Estimation Problem approach following the paper of Drori and Teboulle [4]. Following these works, we know that for a suitable choice of parameters inertial methods ensure fast convergence rates under additional geometry assumptions such as strong convexity. However, the improved convergence results demonstrated in the literature hold only if the function to minimize has a unique minimizer. This extra assumption is limiting, since some common functions (such as the LASSO function or L^1 regularized functions in general) can satisfy some growth condition without having a unique minimizer. The question then arises: is this assumption necessary to prove fast convergence properties?

We propose an approach that aims to avoid this hypothesis while still obtaining fast convergence rates. This strategy allows to extend several known convergence results in the continuous setting (i.e. for dynamical systems associated to inertial schemes). We also provide fast convergence guarantees for the iterates of FISTA (introduced by Beck and Teboulle in [5]) and V-FISTA (proposed by Beck in [6]) in a relaxed setting, showing that this uniqueness assumption is not required for inertial methods to be efficient.

Références

- [1] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR computational mathematics and mathematical physics*, 4.5 (1964): 1-17.
- [2] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Dokl. Akad. Nauk. SSSR*, Vol. 269. No. 3. 1983.
- [3] W. Su, S. Boyd, E. J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17.153 (2016): 1-43.
- [4] Y. Drori, M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145.1-2 (2014): 451-482.
- [5] A. Beck, M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2.1 (2009): 183-202.
- [6] A. Beck. First-order methods in optimization. *Society for Industrial and Applied Mathematics*, 2017.

Exponential Weights Algorithm with Constant Step-Size in Potential Games

- **Maurizio D'Andrea** (Luiss Guido Carli)
- Fabien Gensbittel (Toulouse School of Economics)
- Jérôme Renault (Toulouse School of Economics)

Mots-clé : Online Learning, Nash Equilibrium, Markov Chain, Repeated Games

Résumé : In game theory computing Nash equilibrium is a complex problem, therefore over time the question has arisen as to whether it was possible to construct dynamic processes that converge to this outcome. Furthermore, in recent years, the development of predictive and learning algorithms has garnered increasing interest in the fields of machine learning and artificial intelligence, leading to several applications in everyday life. In this work, we study the long-run behavior of the exponential weights algorithm, introduced for the first time with this name in [3]. Our focus lies on a specific variant where players have access to their pure payoff vectors (that is, including payoffs of unplayed pure strategies). In contrast to [1], where Mertikopoulos et al. explored the problem with a decreasing step-size, we analyze the case with a constant step-size. In the decreasing case they use stochastic approximation methods and the differential equation theory to prove the convergence to strict Nash equilibria in generic potential games; but, the same arguments are not applicable in the constant step size case. Our first result states that if the algorithm converges to a random variable than it is almost surely a weakly stable Nash Equilibrium (*wsne*), where “weakly stable” meaning here that every player’s supported pure strategies yield the same payoff against the supported strategies of every other player. Hence, the existence of a *wsne* is a necessary condition for the convergence of the dynamics. However, this condition is not sufficient, in fact we can show a simple 2-player game with a *wsne* where the algorithm does not converge. Further, we explore the algorithm’s convergence to strict Nash equilibria and the boundary of the game’s mixed strategy space, establishing that it almost surely converges to the boundary in games with a strict Nash equilibrium. Additionally, we refine this result, proving that the algorithm either converges to a strict Nash equilibrium, or it converges to the part of the boundary that does not support any strict Nash equilibria of the game. Then we extend these results to the case of common payoff games. Since the exponential weights algorithm depends only on the difference in payoff where only one strategy is changed, studying a potential game is equivalent to studying a common payoff game where the payoff is the potential function. Finally, we concentrate on "strong coordination games" and establish that within this specific class, the dynamics almost surely converge to a strict Nash equilibrium.

Références

- [1] J. Cohen, A. Héliou, P. Mertikopoulos. Learning with Bandit Feedback in Potential Games. *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA., 2017..
- [2] J. P. Bailey, S. G. Nagarajan, G. Piliouras. Stochastic Multiplicative Weights Updates in Zero-Sum Games. <https://arxiv.org/abs/2110.02134>,, 2021.
- [3] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. *In Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.

Algorithmes inertiels en non convexe (quasar convexe)

- **J. Hermant** (Univ. Bordeaux, Bordeaux INP, CNRS, IMB, Talence, France)
- J.F. Aujol (Univ. Bordeaux, Bordeaux INP, CNRS, IMB, Talence, France)
- C. Dossal (IMT, Univ. Toulouse, INSA Toulouse, Toulouse, France)
- A. Rondepierre (IMT, INSA Toulouse, LAAS, Univ. Toulouse, CNRS, Toulouse, France)

Mots-clé : Optimisation non convexe, Algorithmes inertiels de premier ordre, Optimisation continue et discrète

Résumé : Dans ce travail, nous nous intéressons au comportement des algorithmes d'optimisation inertiels de premier ordre lorsqu'appliqués à une classe de fonctions non convexes appelée fonctions (strongly) quasar convexes. Si il est bien connu que dans le cadre des fonctions convexes des algorithmes de ce type permettent de garantir une convergence plus rapide au minimum de la fonction à minimiser [1], nous savons aussi que des hypothèses de régularité trop faibles empêchent de montrer un bénéfice des algorithmes inertiels [2]. Pour le cas des fonctions smooth (strongly) quasar convexes, il existe des algorithmes de premier ordre permettant d'obtenir une accélération par rapport à la descente de gradient. En particulier, [3] utilise un algorithme inertiel où un paramètre de l'algorithme est calculé via un procédé adaptatif (dichotomie) à chaque itération. Nous pensons que ce procédé adaptatif gêne l'intuition que l'on peut avoir de l'algorithme, et en particulier cela empêche de le voir comme une discrétisation d'un système inertiel continu [4], approche qui dans le cadre convexe s'est montrée très utile pour obtenir et généraliser des résultats de convergence. Ainsi dans ce travail, nous montrons une vitesse de convergence accélérée d'un algorithme inertiel à coefficients explicites dans le cadre de fonctions smooth strongly quasar convexe, au prix d'une restriction sur la courbure. Nous voyons que sans cette restriction, ou dans le cas des fonctions quasar convexes, on peut difficilement éviter un procédé adaptatif de sélection des paramètres. L'algorithme ainsi défini nous permet d'en proposer une analyse continue, dans laquelle nous observons une convergence accélérée de la solution du système inertiel associé sans besoin de restrictions de courbure, ce qui nous permet de mettre en lumière une rupture dans l'analogie discrète et continue dans le cas non convexe. Enfin nous montrons des propriétés plus générales liées à l'optimisation de fonctions non convexes, dont une caractérisation géométrique de ce qui sépare la classe des fonctions PL de celle des fonctions strongly quasar convexes.

Références

- [1] Y.Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$ *Proceedings of the USSR Academy of Sciences*, 543:547–269, 1983.
- [2] Y. Carmon, J. C. Duchi, O. Hinder, A. Sidford Lower Bounds for Finding Stationary Points I *Springer-Verlag*, 1-2–184, 2020.
- [3] O. Hinder, A. Sidford, N. S. Sohoni Near-Optimal Methods for Minimizing Star-Convex Functions and Beyond *Proceedings of Machine Learning Research*, 1894:1938–125, 2020.
- [4] W. Su, S. Boyd, E. J. Candès A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights *Journal of Machine Learning Research*, 1:43–153 vol. 17, 2016.

Numerical Method for a Controlled Sweeping Process with Nonsmooth Sweeping Set

- **Chadi Nour** (Department of Computer Science and Mathematics, Lebanese American University, Byblos Campus, P.O. Box 36 Byblos, Lebanon, cnour@lau.edu.lb)
- Vera Zeidan (Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA, zeidan@msu.edu)

Mots-clé : controlled sweeping process, optimal control, numerical methods, approximations

Résumé : The numerical method developed in [1] for optimal control problems involving sweeping processes with *smooth* sweeping set C is generalized to the case where C is *nonsmooth*, namely, C is the intersection of a finite number of sublevel sets of smooth functions. The novelty of this extension resides in producing for the general setting a different approach, since the one used for the *smooth* sweeping sets is not applicable here. Note that in addition to proving the convergence of our algorithm to an optimal trajectory for the problem, we further confirm the high effectiveness and efficiency of our numerical method by providing an example, for which we calculate, on one hand, an exact optimal solution via the maximum principle of [2], and, on the other hand, a numerical optimal trajectory via our algorithm.

Références

- [1] C. Nour and V. Zeidan. Numerical solution for a controlled nonconvex sweeping process. *IEEE Control Syst. Lett.*, 6:1190–1195, 2022.
- [2] C. Nour and V. Zeidan. Pontryagin-Type Maximum Principle for a Controlled Sweeping Process with Nonsmooth and Unbounded Sweeping Set. *J. Convex Anal.*, 31, 2024 (to appear).

Zero-Sum Stochastic Games with Vanishing Stage Duration and Public Signals

- **Ivan Novikov** (Université Paris-Dauphine, CEREMADE)

Mots-clé : Stochastic games, Zero-sum stochastic games, State-blind stochastic games, Varying stage duration, Viscosity solution, Continuous-time Markov games, Limit value.

Résumé : We consider the behaviour of λ -discounted zero-sum games as the discount factor λ tends to 0 (that is, the players are more and more patient), in the context of games with stage duration. In stochastic games with stage duration h , players act at times $h, 2h, 3h, \dots$, the payoff and the leaving probabilities are proportional to h . When $h \rightarrow 0$, this approximates games in continuous time. The asymptotic behavior of the values (when both λ and h tend to 0) was already studied in the case of stochastic games with perfect observation of the state ([1],[4]). We consider the same question for the case of stochastic games with imperfect observation of the state. In such games, players are given a public signal that depends only on the current state. We have two main results about such games:

- 1) The value of any state-blind stochastic game with stage duration h converges as $h \rightarrow 0$ to a unique viscosity solution of a partial differential equation;
- 2) There exists a stochastic game with public signals, in which there is no limit value (as the discount factor λ goes to 0) if stage duration is 1 ([5]), and we prove that there is a limit value for the same game when stage duration h and discount factor λ both tend to 0. Informally speaking, it means that the limit value in discrete time does not exist, but the limit value in continuous time (i.e., when $h \rightarrow 0$) exists. Such a situation is impossible in the case of stochastic games with perfect observation of the state.

Références

- [1] Abraham Neyman. Stochastic games with short-stage duration. *Dynamic Games and Applications*, 3(2):236–278, 2013.
- [2] Ivan Novikov. Zero-Sum State-Blind Stochastic Games with Vanishing Stage Duration. *Preprint*, 2023.
- [3] Ivan Novikov. Zero-Sum Stochastic Games with Vanishing Stage Duration and Public Signals. *Preprint*, 2023.
- [4] Sylvain Sorin and Guillaume Vigeral. Operator approach to values of stochastic games with varying stage duration. *International Journal of Game Theory*, 45(1):389–410, 2016.
- [5] Bruno Ziliotto. Zero-sum repeated games: Counterexamples to the existence of the asymptotic value and the conjecture $\max_{\min} = \lim v_n$. *The Annals of Probability*, 44(2):1107 – 1133, 2016.

Sur la dualité entre les algorithmes de Frank-Wolfe et les méthodes de plans coupants

- Guilherme Mazanti (L2S, CentraleSupélec, Inria, Université Paris-Saclay)
- **Thibault Moquet** (L2S, CentraleSupélec, Inria, Université Paris-Saclay)
- Laurent Pfeiffer (L2S, CentraleSupélec, Inria, Université Paris-Saclay)

Mots-clé : Optimisation non lisse, Algorithme de Frank-Wolfe, Dualité

Résumé : L'algorithme de Frank-Wolfe [1], ou algorithme du gradient conditionnel, est un algorithme classique d'optimisation convexe. Il permet d'approximer un minimiseur d'une fonction convexe f de classe \mathcal{C}^1 sur un sous-ensemble convexe fermé borné K d'un espace de Hilbert \mathcal{H} , ne nécessitant que d'être capable de minimiser des produits scalaires de la forme $\langle \mu, x \rangle$ pour x dans K , et ce pour n'importe quel μ dans \mathcal{H} .

L'algorithme de plans coupants permet quant à lui de minimiser une fonction convexe et différentiable g sur un sous-ensemble convexe fermé borné D de \mathcal{H} , ne nécessitant que de construire et minimiser une approximation par en-dessous de g qui soit convexe et affine par morceaux.

Je vais vous présenter mes résultats obtenus pendant ma première année de thèse établissant la dualité, au sens de Fenchel-Rockafellar, entre une extension d'une variante plus robuste de l'algorithme de plans coupants, la Level Method [2], et un algorithme pouvant être vu comme une extension de l'algorithme de Frank-Wolfe permettant d'étudier des fonctions s'écrivant comme une somme d'une fonction de classe \mathcal{C}^1 et d'un terme non lisse. Je précise également les garanties théoriques obtenues pour ces méthodes, ainsi qu'un résultat sur leur vitesse de convergence.

Références

- [1] Martin Jaggi. Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. Proceedings of the 30th International Conference on Machine Learning, 427–435, 2013.
- [2] Claude Lemaréchal, Arkadii Nemirovskii, Yurii Nesterov. New variants of bundle methods. *Mathematical programming*, 69:111–147, 1995.

Le jeu derrière la percolation orientée

- Luc Attia (CEREMADE, CNRS, Université Paris Dauphine)
- Raimundo Saona (IST Austria)
- Avelio Sepulveda (CMM, Université du Chili)
- **Bruno Ziliotto** (CEREMADE, CNRS, Université Paris Dauphine)

Mots-clé : Jeu Aléatoire, Percolation, Somme Nulle

Résumé : On considère deux joueurs adversaires qui déplacent un jeton le long des arêtes de \mathbb{Z}^2 , à tour de rôle, l'un dans la direction verticale, l'autre dans la direction horizontale. Le Joueur 1 doit payer un coût aléatoire au Joueur 2 pour chaque arête traversée par le jeton. Ce coût est distribué de manière identique et indépendante, selon une loi de Bernoulli d'un certain paramètre p . Avant que le jeu ne commence, les joueurs sont informés de tous les coûts. Nous étudions la valeur du jeu $v(p)$ où le paiement est la limite inférieure de la moyenne des paiements. Nous montrons d'abord que cette valeur est une constante. Nous montrons également qu'il existe p_1, p_2 dans $(0, 1)$ tels que si $p \leq p_1$, alors $v(p) = 0$, et si $p \geq p_2$, alors $v = 1$. De plus, v est continue en p_2 .

Enfin, nous montrons que p_2 est égal au seuil critique de la percolation orientée classique sur \mathbb{Z}^2 . Ce seuil critique est défini comme l'infimum sur les $p \in [0, 1]$ tel que l'événement "il existe une composante infinie de 1" a une probabilité positive.

Références

- [1] Durrett, R. Oriented percolation in two dimensions. *Annals of Probability*, 12(4), 999-1040.

Zero-sum stochastic games with intermittent observation of the state

- **Guillaume VIGERAL** (Université Paris-Dauphine, CEREMADE)
- Bruno Ziliotto (Université Paris-Dauphine, CEREMADE2)

Mots-clé : stochastic games, public observation, zero-sum

Résumé : We consider asymptotics of finite zero-sum stochastic games as players grow more and more patient. Bewley and Kohlberg[1] established in 1976 that the discounted value converges when players fully observe the state ; in 2016 Ziliotto[2] constructed an example in which the state is not observed at all and the value oscillates. We investigate the middle ground in which the state is publicly observed at some stages, chosen either deterministically or randomly.

Références

- [1] T. Bewley, and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1.3:273–299, 1973.
- [2] B. Ziliotto. Zero-sum repeated games: Counterexamples to the existence of the asymptotic value and the conjecture $\max\min=\lim\min$. *Annals of Probability*, 1107-1133, 2016.

Le chemin des solutions de l'estimateur SLOPE

- **Xavier Dupuis** (IMB, UMR 5584 CNRS, Université de Bourgogne)
- **Patrick Tardivel** (IMB, UMR 5584 CNRS, Université de Bourgogne)

Mots-clé : moindres carrés pénalisés, norme ℓ_1 ordonnée, chemin des solutions, méthode d'homotopie, sélection du paramètre de régularisation

Résumé : L'estimateur SLOPE (Sorted L One Penalized Estimator [1]) est défini comme la solution des moindres carrés pénalisés par une norme ℓ_1 ordonnée :

$$\min_{b \in \mathbb{R}^p} \frac{1}{2} \|y - Xb\|_2^2 + \gamma \sum_{i=1}^p \lambda_i |b|_{\downarrow i} \quad (1)$$

où $|b|_{\downarrow 1} \geq \dots \geq |b|_{\downarrow p} \geq 0$ sont les composantes de b ordonnées par valeur absolue décroissante, $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ sont les paramètres de la norme ℓ_1 ordonnée (fixés, avec $\lambda_1 > 0$) et $\gamma > 0$ est le paramètre de régularisation. Cet estimateur, qui généralise LASSO (pour lequel $\lambda_1 = \dots = \lambda_p = 1$) et OSCAR (pour lequel la suite des λ_i est arithmétique), présente de la parcimonie (composantes nulles) ainsi que de l'appariement (composantes égales en valeur absolue) dont le niveau dépend du paramètre de régularisation γ .

On s'intéresse dans [2] au chemin des solutions de (1) lorsque γ varie dans $(0, +\infty)$. Nous prouvons que ce chemin est continu et linéaire par morceaux, nous caractérisons ses composantes affines et nous proposons un algorithme de type méthode d'homotopie [3] permettant son calcul¹. On note que lorsque le chemin des solutions est connu, la minimisation de la formule SURE ou la minimisation de la somme des carrés résiduelle sur un jeu de données de validation deviennent des problèmes quadratiques par morceaux qui peuvent être résolus exactement, ouvrant la voie à la sélection du paramètre de régularisation γ . Nous montrons par ailleurs sur des jeux de données réels que notre algorithme est compétitif avec les méthodes telles que la descente de coordonnées, les descentes de gradient proximal accélérées ou encore ADMM, qui permettent de résoudre SLOPE à γ fixé mais pas d'en calculer exactement le chemin.

Références

- [1] M. Bogdan, E. van den Berg, C. Sabatti, W. Su, E. Candès. SLOPE — adaptive variable selection via convex optimization. *The Annals of Applied Statistics*, 9(3):1103–1140, 2015.
- [2] X. Dupuis, P. Tardivel. The solution path of SLOPE. hal-04100441, 2023.
- [3] J. Mairal, B. Yu. Complexity analysis of the lasso regularization path. *Proceedings of the 29th International Conference on Machine Learning*, 1835–1842, 2012.

¹le code est disponible sur GitHub : <https://github.com/x-dupuis/slope-path>

Analyzing necessary optimality conditions for a time optimal control problem in Wasserstein spaces.

- **Fernanda Urrea** (INSA Rouen Normandie)
- Hasnaa Zidani (INSA Rouen Normandie)

Mots-clé : Wasserstein spaces, Pontryagin maximum principle, Mean-field optimal control

Résumé : We are interested in developing necessary optimality conditions in the form of a Pontryagin maximum principle for the minimal time Bolza optimal control problem with constraints at the final time. In this problem the dynamic is given by a non-local continuity equation in the Wasserstein space of probability measures, the set of controls are taken in open-loop form and the set of constraints is represented by functional inequalities applied to the terminal time. To this end, review some important results on this domain (see [1, 2]) and we relate the main challenges of this problem from which a fair amount is accounted by the fact that Wasserstein spaces are not Banach and the difficulties on defining differentiability (see [3]).

Références

- [1] Bonnet, Benoît and Frankowska, Hélène Necessary optimality conditions for optimal control problems in Wasserstein spaces. *Applied Mathematics & Optimization*, 84:1281–1330, 2021.
- [2] Bonnet, Benoît and Rossi, Francesco The Pontryagin maximum principle in the Wasserstein space. *Calculus of Variations and Partial Differential Equations*, 58:1–36, 2019.
- [3] Gangbo, Wilfrid and Tudorascu, Adrian. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.

Strict saddle optimization on manifolds

- **Florentin Goyens** (Paris Dauphine - PSL)
- Clément Royer (Paris Dauphine - PSL)

Mots-clé : Riemannian optimization, strict saddle function, second-order method, complexity

Résumé : We consider the minimization of a strict saddle function on a Riemannian manifold, a setting with numerous applications [1]. These functions have the property that the Riemannian hessian has a negative eigenvalue at every critical point that is not a local minimizer. We show that the classical Riemannian trust-region algorithm with exact subproblem minimization finds approximate minimizers in a worst-case number of iterations that improves over known results for nonconvex optimization. The desired accuracy ε only impacts the last few iterations of the algorithm, which benefit from the local quadratic convergence of Newton's method near minimizers. Our complexity bound depends polynomially on landscape parameters which represent negative curvature and strong (geodesic) convexity constants. In that sense, we can say that the worst-case complexity is essentially independent of the accuracy ε . The complexity results are similar to those existing for convex optimization problems. We also present an algorithm with similar complexity guarantees which minimizes the trust-region subproblems approximately with a truncated conjugate gradient scheme. Related works include [2].

Références

- [1] J. Sun, Q. Qing, J. Wright, When are nonconvex problems not scary? arXiv preprint, 2015.
- [2] J. Sun, Q. Qing, J. Wright, A geometric analysis of phase retrieval, Foundations of Computational Mathematics, 2018

Estimateurs à noyaux dans l'étude de la stabilité forte des modèles de risque bidimensionnels

- **Atik Touazi** (University of Algiers, Faculty of Sciences, Algeria)
- Djamil Aissani (University of Bejaia, Research Unit LaMOS, Algeria)
- Smail Adjabi (University of Bejaia, Research Unit LaMOS, Algeria)

Mots-clé : Modélisation stochastique, Modèle de risque , Stabilité forte, Simulation

Résumé : Dans le domaine de l'assurance, la modélisation stochastique est essentielle pour comprendre l'évolution temporelle des risques. Les processus stochastiques sont souvent utilisés pour modéliser les sinistres dans le temps, prenant en compte les événements aléatoires qui peuvent affecter la fréquence et la gravité des réclamations. Cela permet d'estimer les réserves nécessaires pour couvrir ces sinistres futurs. la théorie de la ruine offre un cadre analytique pour évaluer les risques financiers auxquels sont confrontées les entreprises d'assurance. Ces approches permettent de mieux comprendre la probabilité de ruine et d'ajuster les stratégies de gestion des risques pour garantir la stabilité financière à long terme.

Pour une étude pratique, différentes lois de probabilité peuvent être utilisées pour modéliser le nombre et le montant des réclamations. En réalité, la détermination de ces lois de probabilité ne peut être obtenue qu'à partir d'un échantillon d'observations. Deux types de perturbations peuvent être pris en compte dans l'analyse de la stabilité forte. La première est due aux données aberrantes, qui peuvent avoir de multiples causes. La deuxième perturbation est due aux erreurs résultant de l'utilisation des techniques d'estimation fonctionnelle pour la détermination des lois de probabilités des paramètres gouvernant le modèle.

En utilisant l'approche de stabilité forte, des bornes de stabilité des probabilités de ruine sont établies par Kalashnikov (see [1]). Par la suite, plusieurs travaux pour les modèles unidimensionnels ont été réalisés (see [2, 3, 4]).

L'objectif de ce travail est de démontrer l'applicabilité de la méthode de stabilité forte à l'étude d'un modèle de risque à deux dimensions lorsque la loi des montants de réclamation ou la loi des inter-sinistres est générale et inconnue. Dans ce cas, on fait appel à la méthode du noyau pour estimer la fonction densité de la distribution considérée. Par la suite, en utilisant l'approche de simulation, nous évaluons numériquement la borne de stabilité forte (erreur d'approximation) entre les probabilités de ruine des deux modèles de risque, idéal et perturbé.

Références

- [1] V. Kalashnikov. The stability concept for stochastic risk models. *Working Paper Nr 166, Laboratory of Actuarial Mathematics. University of Copenhagen*, 2000.
- [2] A. Touazi, Z. Benouaret, D. Aissani and S. Adjabi. Nonparametric estimation of the claim amount in the strong stability analysis of the classical risk model. *Insurance : Mathematics and Economics*, 74:78–83, 2017.
- [3] A. Bareche, M. Cherfaoui. Sensitivity of the stability bound for ruin probabilities to claim distributions. *Methodology and computing in applied probability*, 21:1259–1281, 2019.
- [4] Z. Harfouche, A. Bareche. Semi-parametric approach for approximating the ruin probability of classical risk models with large claims. *Communication in Statistics- Simulation and Computation*, doi:10.1080/03610918.2021.1992636, 2021.

A Lagrange-Galerkin scheme for first order mean field games systems

- Elisabetta Carlini (Sapienza Università di Roma)
- Francisco J. Silva (Université de Limoges)
- **Ahmad Zorkot** (Université de Limoges)

Mots-clé : Mean field games, optimal control, semi Lagrangian scheme, Lagrange-Galerkin scheme, numerical analysis

Résumé : In this work, we consider a first order mean field games system with non-local couplings. A Lagrange-Galerkin scheme for the continuity equation, coupled with a semi-Lagrangian scheme for the Hamilton-Jacobi-Bellman equation, is proposed to discretize the mean field games system. The convergence of solutions to the scheme towards a solution to the mean field game system is established in arbitrary space dimensions. The scheme is implemented to approximate two mean field games systems in dimension one and two.

Références

- [1] E Carlini, Francisco José Silva and Ahmad Zorkot. A Lagrange-Galerkin scheme for first order mean field games systems *SIAM J. Numer. Anal.* 2023.

Two Optimal Transportation approaches to measuring Risk.

- Olivier P. FAUGERAS (Toulouse School of Economics, University of Toulouse Capitole)

Mots-clé : Risk measure, Optimal Transportation, Hemi-metric, Optimal quantization

Résumé : The evaluation and comparison of risks are basic tasks of risk analysis in Insurance and Finance. In this talk, we show how Optimal Transportation can be employed for this purpose.

The first approach, based on [1], considers that risk is a relative notion between two distributions, quantifying an order relation. It introduces the notion of risk excess measure of one risk distribution Q w.r.t. a benchmark risk P . More precisely, it measures the risk excess of Q over P by a hemi-metric $D_+(Q, P)$ on the space of probability measures. $D_+(Q, P)$ is a “one-sided distance encoding an order” on the space $(\mathcal{M}_1(E), \prec)$ of probability measures, where \prec is a given stochastic (pre)order. The stochastic order \prec is related to the ordering \leq on the underlying space E . This allows to consider for a quantitative one-sided comparison of risks at the level of probability measures as an extension of the order and distance structure on E . We discuss several classes of risk excess measures $D_+(Q, P)$ and consider the question when these are given as order extensions of hemi-distances d_+ on the underlying space E . Several relevant hemi-distances are induced by mass transportation and by function class induced orderings, thus giving access to natural interpretation. One particular extension is given by a version of the Kantorovich–Rubinstein theorem for hemi-distances. Our view towards measuring risk excess adds to the usually considered method to compare risks of Q and P by the values $\rho(Q)$, $\rho(P)$ of a risk measure ρ . We argue that the difference $\rho(Q) - \rho(P)$ neglects relevant aspects of the risk excess which are adequately described by the new notion of risk excess measure.

The second approach, based on [2], proposes another novel approach in the assessment of a random risk variable X by introducing magnitude-propensity risk measures (m_X, p_X) . This bivariate measure intends to account for the dual aspect of risk, where the magnitudes x of X tell how high are the losses incurred, whereas the probabilities (propensities) $P(X = x)$ reveal how often one has to expect to suffer such losses. The basic idea is to simultaneously quantify both the severity m_X and the propensity p_X of the real-valued risk X . This is to be contrasted with traditional univariate risk measures, like VaR or CVaR, which typically conflate both effects. In its simplest form, (m_X, p_X) is obtained by mass transportation in Wasserstein metric of the law of X to a two-points $\{0, m_X\}$ discrete distribution with mass p_X at m_X . The approach can also be formulated as a constrained optimal quantization problem. This allows for an informative comparison of risks on both the magnitude and propensity scales. Several examples illustrate the usefulness of the proposed approach. Some variants, (multivariate) extensions and applications (portfolio analysis) are also considered.

Références

- [1] O. P. Faugeras, Rüschendorf, L. Risk excess measures induced by hemi-metrics. *Probability, Uncertainty and Quantitative Risk*, Vol. 3, 6, 2018. <https://doi.org/10.1186/s41546-018-0032-0>
- [2] O. P. Faugeras, G. Pagès. Risk Quantization by Magnitude and Propensity. hal-03233068v2, Submitted, in revision, 2021.

Optimal control problems for multi-agent systems: analysis from a Lagrangian and an Eulerian perspective

- **Giulia Cavagnari** (Politecnico di Milano)
- Stefano Lisini (University of Pavia)
- Carlo Orrieri (University of Pavia)
- Giuseppe Savaré (Bocconi University)

Mots-clé : Mean-field optimal control, Superposition principle, Gamma-convergence

Résumé : The study of evolution equations in the Wasserstein space is getting an increasing attention as a model for multi-agent dynamics.

We will discuss different approaches for modeling optimal control problems for interacting multi-particle systems: an Eulerian one in the metric Wasserstein space of probability measures; a Lagrangian one (à la Lions) in the Hilbert space $L^2_{\mathbb{P}}(\Omega)$, where (Ω, \mathbb{P}) is a given space of parametrizations; and a so-called Kantorovich formulation which serves as an intermediate point of view and it is based on the Superposition Principle of Ambrosio-Gigli-Savaré. Requiring that the Eulerian evolution is the time-dependent law of the Lagrangian process, we prove the equivalence of the value functions in the three formulations, highlighting the pros/cons of them. These results turn out to be interesting also in light of recent studies (Jimenez-Marigonda-Quincampoix, Jimenez) on the relations between different definitions of viscosity solutions for Hamilton-Jacobi equations in Wasserstein spaces and in the lifted space $L^2_{\mathbb{P}}(\Omega)$.

Finally, we derive consistency results as the number of particles/agents tends to infinity.

This is based on a joint paper with Stefano Lisini, Carlo Orrieri (University of Pavia) and Giuseppe Savaré (Bocconi University).

Mathematical perspectives on Transformers

- **Borjan Geshkovski** (INRIA & Sorbonne Université)
- **Cyril Letrouit** (CNRS & Université Paris Saclay)
- Yury Polyanskiy (Massachusetts Institute of Technology)
- Philippe Rigollet (Massachusetts Institute of Technology)

Mots-clé : Machine learning, Interacting particle systems, Clustering

Résumé : This talk will report on several results, insights and perspectives we have found regarding *Transformers*, a neural network architecture that plays a central role in the inner workings of large language models. We model Transformers as interacting particle systems (each particle representing a token), governed by a non-linear coupling called *self-attention*. When considering pure self-attention Transformers, we show that trained representations cluster in long time to different geometric configurations determined by spectral properties of the model weights. We also cover Transformers with layer normalization, which amounts to considering the interacting particle system on the sphere. On high-dimensional spheres, we prove that all randomly initialized particles converge to a single cluster. The result is made more precise by describing the precise phase transition between the clustering and non-clustering regimes. The appearance of metastability, and ideas for the low-dimensional regime, will be discussed.

Références

- [1] B. Geshkovski, C. Letrouit, Y. Polyanskiy and P. Rigollet. The emergence of clusters in self-attention dynamics. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [2] B. Geshkovski, C. Letrouit, Y. Polyanskiy and P. Rigollet. A mathematical perspective on Transformers. *preprint*, 2023.

Approximations de la fonction de transport optimal par des flux de systèmes linéaires en les contrôles

- **Alessandro Scagliotti** (TU Munich & Munich Center for Machine Learning, Munich, Allemagne)
- Sara Farinelli (Lagrange Mathematics and Computation Research Center, Paris, France)

Mots-clé : Γ -convergence, Transport Optimal, Contrôle Optimal

Résumé : Dans cette présentation, on étudie le problème de la reconstruction de la fonction de transport optimal T entre deux mesures absolument continues $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, et, pour l'approximations, on utilise des flux des systèmes dynamiques en \mathbb{R}^n , linéaires en les contrôles. Nous montrons d'abord que, sous des hypothèses appropriées sur μ, ν et sur les champs de vecteurs contrôlés, la fonction de transport optimal est contenue dans la C_c^0 -clôture des flux générés par le système.

En supposant que des approximations discrètes μ_N, ν_N des mesures originales μ, ν soient connues, on utilise un plan optimal discret γ_N pour définir un problème de contrôle optimal. Avec un argument de Γ -convergence, on prouve que ses solutions correspondent à des flux qui approchent la fonction de transport optimal T .

Enfin, en vertu du Principe du Maximum de Pontryagin, on propose un schéma numérique itératif pour la résolution du problème de contrôle optimal, ce qui donne lieu à un algorithme pour le calcul pratique de l'approximation de la fonction de transport optimal. Cette approche peut être interprétée comme la construction d'un "Normalizing Flow" au moyen d'un réseau neuronal résiduel (ResNet).

Références

- [1] A. Scagliotti, S. Farinelli. Normalizing flows as approximations of optimal transport maps via linear-control neural ODEs. *arXiv preprint*, arXiv:2311.01404v2, 2023.

The gradient's limit of a definable family of functions is a conservative set-valued field

- **Sholom Schechtman** (SAMOVAR, Télécom Sudparis, Institut Polytechnique de Paris)

Mots-clé : Conservative set-valued fields, Clarke subgradients, smoothing methods

Résumé : In this work, given a family of real-valued functions $(f_a)_{a>0}$ that converges, when $a \rightarrow 0$, to some function $F : \mathbb{R}^d \rightarrow \mathbb{R}$ we are interested in the behavior of the limit of the corresponding family of (sub)-gradients $(\partial f_a)_{a>0}$. Formally, we are interested in properties that satisfies:

$$D_F(x) := \{v \in \mathbb{R}^d : \text{there is } (x_n, v_n, a_n) \rightarrow (x, v, 0) \text{ such that } a_n \searrow 0 \text{ and } v_n \in \partial f_a(x_n)\}.$$

Ideally, one would wish that D_F reduces to one of the common first-order operators: to the gradient if F is differentiable, to the convex subgradient if it is convex or to the Frechet's or Clarke's subgradient if F is merely locally Lipschitz continuous. If the family (f_a) is convex, then this is indeed the case thanks to the celebrated Attouch's theorem ([1]). Unfortunately, it is well-known that without assumptions alike to some (approximate) convexity such a result does not hold even if every f_a and F are smooth.

Main result. Our main result states that if the function $x, a \mapsto f_a(x)$ is definable in an o-minimal structure (for instance, is semialgebraic, subanalytic, or any composition of the latter), then D_F is a **conservative set-valued field**: a notion introduced by Bolte and Pauwels in the context of automatic differentiation ([3]).

While such a result is interesting from purely theoretical point of view, the operator D_F naturally appears in the analysis of smoothing methods. In these, (f_a) represents a family of smooth approximations of a nonsmooth, nonconvex F , which we optimize by slowly reducing the parameter a ([2]). In this setting, a smoothing method will ususally find a point x^* that is D_F -critical: $0 \in D_F(x^*)$. Thus, the characterization of D_F as a conservative set-valued field is important from both theoretical and practical perspective.

In particular, our result shows that while D_F might be strictly larger than the Clarke's subgradient it enjoys many of desirable properties of a first-order operator. First, for almost every $x \in \mathbb{R}^d$, it holds that $D_F(x) = \nabla F(x)$ (gradient consistency in the sense of [2]). Second, since F is definable, its domain can be partitioned into manifolds (\mathcal{M}_i) such that on each of \mathcal{M}_i , F is differentiable and for $x \in \mathcal{M}_i$ the projection of $D_F(x)$ onto the tangent plane of \mathcal{M}_i at x is just the Riemannian gradient of $F|_{\mathcal{M}_i}$. Lastly, it always holds that $\partial F(x) \subset \text{conv} D_F(x)$ (see [3]).

Références

- [1] H. Attouch Convergence de fonctionnelles convexes *Journées d'Analyse Non Linéaire: Proceedings, Besançon, France, June 1977*, Springer, 1–40, 2006
- [2] X. Chen Smoothing methods for nonsmooth, nonconvex minimization *Mathematical Programming*, 134:71–99, 2012
- [3] J. Bolte, E. Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning *Mathematical Programming*, 188:19–51, 2021

Approximation et perturbation de solutions stables d'un système MFG du second ordre stationnaire

- Jules Berry (IRMAR – INSA Rennes)
- Olivier Ley (IRMAR – INSA Rennes)
- Francisco Silva (XLIM – Université de Limoges)

Mots-clé : Jeux à champ moyen, Éléments finis, Méthode de Newton

Résumé : L'analyse numérique des solutions des systèmes de jeux à champ moyen (MFG) a fait l'objet de nombreux travaux au cours des quinze dernières années mais le cadre généralement adopté est celui de la monotonie introduit par Lasry et Lions. En effet l'hypothèse de monotonie de Lasry-Lions est connue pour être une condition suffisante pour l'unicité des solutions des systèmes MFG. Pour une certaine famille de jeux à champ moyen une notion de solution stable a été proposée par Briani-Cardaliaguet dans [1] dans le cadre du système MFG en horizon de temps fini du second ordre. Celles-ci ne sont en général pas uniques en l'absence de l'hypothèse de monotonie mais présentent certaines bonnes propriétés. Par exemple il est démontré dans [1] que les solutions stables sont isolées.

Nous introduisons un cadre général pour l'étude des approximations numériques des solutions stables des systèmes de jeux à champ moyen. Pour illustrer la démarche nous nous focalisons sur un exemple très simple de système MFG stationnaire du second ordre avec couplage local et Hamiltonien quadratique de la forme

$$\begin{cases} -\Delta u + \frac{1}{2}|Du|^2 + \lambda u = f(m) & \text{sur } \mathbb{T}^d, \\ -\Delta m - \operatorname{div}(mDu) + \lambda m = \lambda m_0 & \text{sur } \mathbb{T}^d, \end{cases} \quad (1)$$

où $\lambda > 0$. Nous donnons des conditions suffisantes pour la stabilité des solutions du système (1). Nous ré-exprimons ensuite les solutions du système (1) comme des zéros d'une certaine application non-linéaire F et nous établissons le fait que si (u, m) est une solution stable de (1) alors la différentielle de F en (u, m) est un isomorphisme. Ce fait est ensuite exploité pour étudier des perturbations du système (1), l'approximation des solutions stables de (1) par éléments finis et la convergence de la méthode de Newton appliquée à celles-ci.

Références

- [1] A. Briani and P. Cardaliaguet. Stable solutions in potential mean field game systems *NoDEA*, 2018.

Optimal structures for a model of crop irrigation with constraints

- **Ruben Chenevat** (INRAE - UMR MISTEA)
- Alain Rapaport (INRAE - UMR MISTEA)
- Bruno Cheviron (INRAE - UMR G-EAU)
- Sébastien Roux (INRAE - UMR MISTEA)

Mots-clé : Optimal control, Non-autonomous dynamics, Non-smooth dynamics, State constraints, Singular arcs, Water management

Résumé : Dans un contexte actuel où la population mondiale continue de croître et les tensions climatiques s'intensifient, la question de la gestion des ressources naturelles se pose. La hausse des températures, l'augmentation de l'évapotranspiration des sols, l'assèchement des nappes phréatiques, ou la multiplication d'événements météorologiques extrêmes et difficiles à prévoir sont autant de facteurs impactant les pratiques agricoles. Afin de répondre à la demande de production croissante tout en s'adaptant à ces nouvelles pressions, il existe différents leviers d'action. En particulier l'adoption d'outils de pilotage de l'irrigation offre des bons potentiels d'économie d'eau face à un contexte de déficit hydrique (voir [1]).

Nous nous intéressons alors à un modèle simple de culture (voir [2]), décrivant les phénomènes physiques importants de la croissance de plante, et dont l'écriture est adaptée à une étude analytique via la théorie du contrôle optimal. Nous proposons d'étudier ce problème d'optimisation sous une formulation permettant de traiter différents types de critères et de cibles (selon les objectifs agronomiques choisis : production de biomasse, économie d'eau, balance financière). De plus, nous cherchons à déterminer à la fois les meilleures et les pires stratégies, afin de quantifier les gains potentiels résultant d'une bonne pratique.

Les caractéristiques du système apportent un certain nombre de difficultés mathématiques à surmonter dans le but de résoudre analytiquement le problème d'optimisation. Nous allons notamment gérer les contraintes d'état avec une technique inspirée de [3] (paragraphe 10), où une contrainte est traduite en une équation et l'introduction d'une nouvelle variable d'état. Nous démontrons ensuite en appliquant le principe du maximum de Pontryagin que les stratégies optimales peuvent s'écrire sous forme de feedback, et qu'elles sont décrites grâce à des structures entièrement déterminées par 1 ou 2 paramètres. La résolution numérique de ce sous-problème d'optimisation en 1 ou 2 dimensions est efficace et permet ainsi de tester, d'évaluer, de valider le modèle sur une large gamme de données d'entrée, en vue enfin de contribuer concrètement au développement d'outils de pilotage de l'irrigation.

Références

- [1] Serra-Witling C., Molle B., Cheviron B. Plot level assessment of irrigation water savings due to the shift from sprinkler to localized irrigation systems or to the use of soil hydric status probes. *Agricultural Water Management*, 223 : 105682, 2019.
- [2] K. Boumaza, N. Kalboussi, A. Rapaport, S. Roux, C. Sinfort Optimal control of a crop irrigation model under water scarcity. *Optimal Control Applications and Methods*, 42(6) 1612-1631, 2021.
- [3] R. F. Hartl, S. P. Sethi, R. G. Vickson A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, Vol. 37, No. 2, pp. 181-218, 1995.

Quantitative Stability of the Pushforward Operation by an Optimal Transport Map

- Guillaume Carlier (Université Paris-Dauphine & Inria)
- Alex Delalande (EPFL)
- Quentin Mérigot (Université Paris-Saclay)

Mots-clé : Optimal transport, Pushforward measure, Singularities of convex functions.

Résumé : We study the quantitative stability of the mapping that to a measure associates its pushforward measure by a fixed (non-smooth) optimal transport map. We exhibit a tight Hölder-behavior for this operation under minimal assumptions. Our proof essentially relies on a new bound that quantifies the size of the singular sets of a convex and Lipschitz continuous function on a bounded domain.

Loss control regions in optimal control problems

- T erence Bayen (Avignon Universit e. Laboratoire de Math ematiques d'Avignon. EA 2151)
- **Anas Bouali** (Avignon Universit e. Laboratoire de Math ematiques d'Avignon. EA 2151)
- Lo ic Bourdin (Institut de recherche XLIM. UMR CNRS 7252. Universit e de Limoges)
- Olivier Cots (Toulouse INP. Institut National Polytechnique)

Keywords: Optimal control, loss control regions, non-permanent controls, direct methods, shooting methods.

In this presentation, we address optimal control problems involving loss control regions. In this context, the state space is partitioned into disjoint sets referred to as regions, which are classified into two types: *control regions* and *loss control regions*. When the state belongs to a control region, the control is permanent (i.e., the control value can be modified at any time). On the other hand, when the state belongs to a loss control region, the control must remain constant, equal to the last assigned value before the state enters the loss control region, and this value is kept until the state exits this region. The goal of this presentation is twofold. First, we derive a corresponding Pontryagin maximum principle. Second, we propose a two-step numerical scheme to solve optimal control problems with loss control regions. The approach is based on a direct numerical method applied to a regularized problem, which initializes an indirect numerical method based on the previously mentioned optimality conditions and applied to the original problem. Lastly, we apply this numerical approach to several illustrative examples.

R ef erences

- [1] Loss control regions in optimal control problems. *with T erence Bayen, Lo ic Bourdin and Olivier Cots.* (submitted) <https://hal-univ-avignon.archives-ouvertes.fr/hal-04137550>
- [2] The hybrid maximum principle for optimal control problems with spatially heterogeneous dynamics is a consequence of a Pontryagin maximum principle for L^1_{\square} -local solutions. *with T erence Bayen and Lo ic Bourdin.* (in revision) <https://hal-univ-avignon.archives-ouvertes.fr/hal-03985420>
- [3] Hybrid maximum principle with regionally switching parameter. *with T erence Bayen and Lo ic Bourdin.* (in revision) <https://hal.archives-ouvertes.fr/hal-03638701>

Analyse de sensibilité et contrôle optimal pour la Chemical Master Equation

- Frédéric Bonnans (L2S CentraleSupélec & INRIA Saclay-Île-de-France)
- **Arthur Bottois** (L2S CentraleSupélec & INRIA Saclay-Île-de-France)
- Laurent Pfeiffer (L2S CentraleSupélec & INRIA Saclay-Île-de-France)

Mots-clé : Chemical Master Equation, chaînes de Markov en temps continu, contrôle optimal

Résumé :

On considère un système chimique composé de d espèces X_1, \dots, X_d interagissant via R réactions,

$$\text{pour } 1 \leq r \leq R, \quad R_r : \alpha_r \cdot X \xrightarrow{a_r(t,x)} \beta_r \cdot X, \quad \text{avec } X = (X_1, \dots, X_d).$$

Chaque réaction R_r est caractérisée par ses coefficients stœchiométriques $\alpha_r, \beta_r \in \mathbb{N}^d$ et son taux de réaction $a_r(t, x)$. Ce taux de réaction peut dépendre du temps $t \geq 0$ et de l'état $x \in \mathbb{N}^d$ du système. Être dans l'état $x = (x_1, \dots, x_d)$ signifie qu'il y a x_i molécules de l'espèce X_i , pour tout $1 \leq i \leq d$. Un tel système chimique peut être décrit par une chaîne de Markov en temps continu $X(t)$, de loi $p(t, x) = \mathbb{P}[X(t) = x]$, la probabilité d'être dans l'état x au temps t . Cette loi vérifie une équation de Kolmogorov, appelée ici Chemical Master Equation (CME),

$$\frac{\partial p}{\partial t}(t, x) = \sum_{r=1}^R \left(a_r(t, x - \nu_r) p(t, x - \nu_r) - a_r(t, x) p(t, x) \right), \quad \text{avec } \nu_r = \beta_r - \alpha_r. \quad (\text{CME})$$

Cette équation est utilisée en biologie (cf. [1]) pour modéliser des réactions chimiques ayant lieu dans des cellules, afin de prendre en compte le caractère aléatoire de ces réactions et dans un contexte où les effectifs des molécules mises en jeu sont trop faibles pour passer au continu sur l'espace d'état $\mathcal{S} = \mathbb{N}^d$. La CME est donc une collection dénombrable d'équations différentielles ordinaires.

Dans le cas simple où les taux de réaction a_r sont bornés, le théorème de Cauchy-Lipschitz assure l'existence et l'unicité de la solution, tandis que l'application du théorème des fonctions implicites permet de mener une analyse de sensibilité de cette solution vis-à-vis des données du problème (condition initiale, taux de réaction, ...). Cependant, les applications en biologie amènent plutôt à considérer des taux de réaction proportionnels au nombre de réactifs et donc non bornés. Dans ce cadre, l'unicité de la solution n'est plus assurée. On peut néanmoins s'appuyer sur la notion pertinente (au point de vue physique) de solution minimale (cf. [2]).

En utilisant des estimations de stabilité pour la solution minimale, on établit la différentiabilité de cette solution par rapport aux données du problème. On donne ensuite une formule de dérivée directionnelle faisant intervenir une équation adjointe, pour laquelle on définit également un concept de solution minimale. Cela nous permet de démontrer des conditions d'optimalité (sous la forme d'un principe de Pontryagin) pour une classe de problèmes d'optimisation pour lesquels on contrôle les vitesses de réaction.

Références

- [1] S. Winkelmann & C. Schütte. Stochastic dynamics in computational biology. *Vol. 645, Berlin/Heidelberg, Germany: Springer, 2020.*
- [2] L. Gauckler & H. Yserentant. Regularity and approximability of the solutions to the chemical master equation. *ESAIM: M2AN*, 48–6:1757–1775, 2014.

Fixed points of regular set-valued mappings in quasi-metric spaces

- Huu Tron Nguyen (Department of Mathematics and Statistics, Quy Nhon University)
- Michel Théra (Mathematics and Computer Science Department, University of Limoges)

Mots-clé : Regularity, Pseudo-Lipschitzness contraction, orbital regularity, coincidence point, fixed point, cyclic fixed point, quasi-metric space

Résumé : This talk focuses on the existence of fixed points for set-valued mappings defined on quasi-metric spaces complete or non-complete. Quasi-metrics play a central role in the theory of spaces of homogeneous type and have numerous recent applications both in pure and applied mathematics, for example, in the questions of existence and uniqueness of Hamilton-acobi equations, in rate-independent models for plasticity, shape-memory alloys, models for material failure, automated taxonomy construction. In the context mentioned above, we use the notion of *orbital regularity/ orbital pseudo-Lipschitzness* of set-valued mappings defined by Ioffe to obtain the estimate of the distance from a given point to the set of fixed points of the mapping under consideration. Existence of double fixed points, coincidence points of pairs of set-valued mappings from a quasi-metric space into another one will also be given. Moreover, the existence of cyclic fixed points for a collection of set-valued mappings and estimations of the distance from a given point to the set of fixed points or the distance between two sets of fixed points will be also provided. Our results generalize recent results given by Ait Mansour et al. [1], Dontchev & Rockafellar [4], Ioffe [2], Lim [3] and Nguyen Huu Tron [5].

Références

- [1] Ait Mansour, M., Bahraoui, M.A., El Bekkali, A. Metric Regularity and Lyusternik-Graves Theorem via Approximate Fixed Points of Set-Valued Maps in Noncomplete Metric Spaces. *Set-Valued Var. Anal.* 30:233-256, 2022.
- [2] Ioffe, A.D. Variational Analysis of Regular Mappings: Theory and Applications. 95. Springer Monographs in Mathematics, Springer, 2017.
- [3] Lim, T.-C. On fixed-point stability for set-valued contractive mappings with applications to generalized differential equations. *J. Math. Anal. Appl.* 110: 436-441, 1985.
- [4] Dontchev, A.L., Rockafellar, R.T. Implicit Functions and Solution Mappings: A View from Variational Analysis. 2nd Edition, 466: Springer, New York 2014.
- [5] Tron, N.H. Coincidence and Fixed Points of Set-Valued Mappings Via Regularity in Metric Spaces. *Set-Valued Var. Anal.* 31: 2, Paper No. 17, 2023.

Optimization Landscape of Matrix Square Root Approximation

- Alexandre Allauzen (ESPCI-PSL)
- **Iskander Legheraba** (Université Paris Dauphine-PSL)
- Clément W. Royer (Université Paris Dauphine-PSL)

Mots-clé : Numerical optimization, eigenvalues, singular values, numerical linear algebra

Résumé : The efficacy of deep neural networks has spurred extensive research into their training processes. Landscape analysis, emphasizing loss function critical points, has been particularly successful in studying linear networks. Loss landscape exploration has also been conducted for continuous-in-depth architectures inspired by NeuralODE, a recently introduced architecture based on ordinary differential equations. However, this analysis does not apply to practical NeuralODE implementations.

In this talk, we focus on a discretized NeuralODE implementation obtained using the forward Euler method, and analyze the landscape of its training process. Assuming whitened training data and a symmetric positive definite target matrix, we classify global minima and identify two strict saddle point classes, including a novel class absent in linear networks. A full landscape description is provided for a scalar target matrix and we show that in that case the loss landscape satisfies the strict saddle property.

Optimal transport-based robust optimization with SkWDR0

- Vincent Florian (Inria Grenoble, LJK)

Mots-clé : machine learning, optimal transport, robust optimization, Scikit-Learn, PyTorch

Résumé : Optimal transport-based neighborhoods are popular to model data uncertainty in robust optimization [4]. Dealing with Wasserstein distances may however raise numerical issues. Inspired by the entropy regularization in optimal transport [3], a promising smoothing approach was developed by [1] to solve distributionally robust problems.

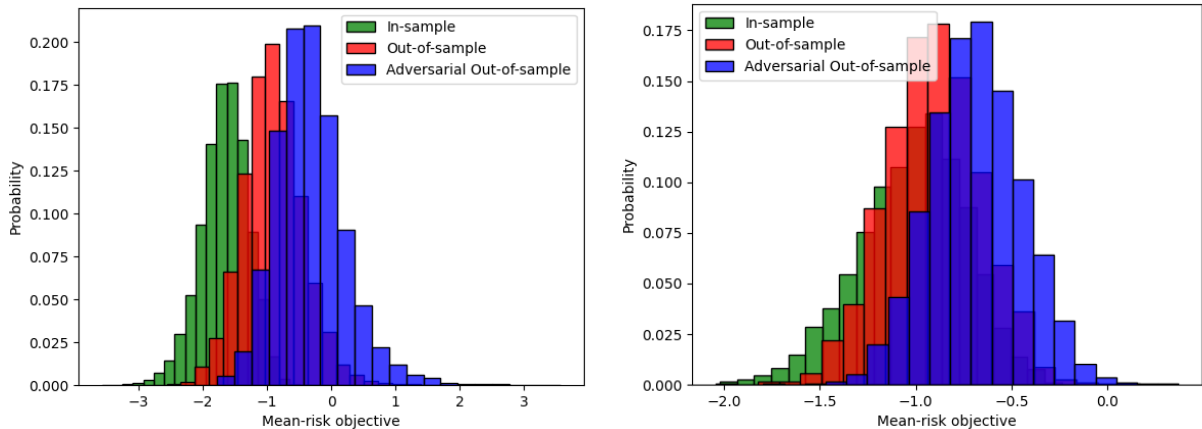


Figure 1: Illustration on which we plot the distributions of losses, with and without robustness. In green, the loss observed in the training samples; in red the loss for new samples; and in blue the loss for an adversarial perturbation of the data. Observe how the robust solution is more faithful to its training promises after testing, and resists adversarial perturbations better.

This communication is about `SkWDR0`, a Python toolbox for entropy-regularized robust models. `SkWDR0` proposes easy-to-use interfaces with two popular machine learning libraries: `Scikit-Learn` to obtain robust counterparts of standard models, and `PyTorch` to wrap deep learning models with only a couple of additional lines of code. Finally, `SkWDR0` uses an automatic calibration of hyperparameters, with heuristics from [1].

Références

- [1] W. Azizian, F. Iutzeler, and J. Malick. Exact Generalization Guarantees for (Regularized) Wasserstein Distributionally Robust Models. *In NeurIPS 2023*.
- [2] W. Azizian, F. Iutzeler, and J. Malick. Regularization for wasserstein distributionally robust optimization. *ESAIM: COCV, 2023*.
- [3] G. Peyré and M. Cuturi. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends in Machine Learning, 2019*
- [4] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations *Mathematical Programming, 171, 2018*

Incomplete markets with a countable number of states: Equilibrium and No-Arbitrage

- **Jean-Marc Bonnisseau** (Université Paris 1 Panthéon-Sorbonne, Centre d'économie de la Sorbonne, Paris School of Economics)
- Cuong Le Van (IPAG Business School, TIMAS, Paris School of Economics)
- Cuong Tran Viet (Université Paris 1 Panthéon-Sorbonne, CES)

Mots-clé : Financial Equilibrium, Financial Markets, Infinite State Space, Incomplete Market, Equilibrium, Exchange Economy, Assets, Asset Market Equilibrium, No-Arbitrage Prices, No-Arbitrage Condition, Truncation

Résumé : In this paper, we prove the existence of an equilibrium in a two-period model à la Hart with incomplete markets and a countable number of states. Moreover, under some restrictions on the returns matrix, an equilibrium asset price is a no-arbitrage price. Conversely, we consider a sequence of equilibria corresponding to an increasing number of states associated with a given no-arbitrage asset price. If the sequence of commodity prices has a non-zero cluster point for the product topology, then the limits of these prices and of the allocations (assets, commodities) constitute, together with the given asset price, an equilibrium with a countable number of states.

Sur l'apprentissage de combinaisons linéaires par programmation convexe sur les mesures.

- Joël Garde (Télécom Paris — LTCI)
- Olivier Fercoq (Télécom Paris — LTCI)
- Hicham Janati (Télécom Paris — LTCI)

Mots-clé : Beurling Lasso, Problèmes Inverses, Parcimonie, Variation totale, hors-grille

Résumé : On s'intéresse aux combinaisons linéaires comme solution de problèmes inverses et leur paramétrisation par des mesures discrètes, permettant de les résoudre par programmation convexe. On emploiera pour l'apprentissage le Beurling Lasso [3], qui solutionne les moindres carrés régularisés par la variation totale sur un espace de mesures.

Le Beurling Lasso s'analyse comme une formulation hors grille du lasso, évitant la discrétisation de l'espace des paramètres. C'est une relaxation continue de l'algorithme du lasso: il permet de convexifier le problème en échange du caractère finit de la dimension ambiante. La dimension infinie va alors rendre la résolution algorithmique difficile.

Cette difficulté est surmontée par l'approche de Frank–Wolfe grâce à un oracle qui ajoute successivement des particules à l'itéré [1]. La structure hors-grille permet de prendre en compte la géométrie de l'espace des paramètres et de déplacer les particules [4]. L'algorithme du Conic Particle Gradient Descent [2] embrasse cette structure, en simulant un flot de gradient sur l'espace des mesures par une descente de gradient sur un nuage de particules. La géométrie de l'espace des paramètres est alors propagée aux mesures par la métrique de Wasserstein.

On propose ici de décrire ces approches dans le cadre de la localisation de sources en imagerie cérébrale à partir de mesures électro-encéphalographiques. Cette nouvelle application introduit pour la première fois des paramètres en dimension trois avec des mesures à valeur vectorielle pour représenter la position et l'amplitude des sources. La difficulté de la tâche est renforcée par le noyau issu de la solution du problème direct d'électro-magnétisme, plus complexe que les noyaux Gaussiens ou de Dirichlet utilisés habituellement dans les exemples de la littérature. On s'attache à obtenir un algorithme performant, notamment avec une nouvelle approche dite variationnelle [5], qui souligne le caractère séparable du problème par un traitement différencié des amplitudes et des positions.

Références

- [1] Bredies, K. & Pikkarainen, H. Inverse Problems in Spaces of Measures. *ESAIM: Control, Optimisation And Calculus Of Variations*. **19**, 190-218 (2013,1)
- [2] Chizat, L. Sparse Optimization on Measures with Over-Parameterized Gradient Descent. *Mathematical Programming*. **194**, 487-532 (2022,7,1)
- [3] Castro, Y. & Gamboa, F. Exact Reconstruction Using Beurling Minimal Extrapolation. *Journal Of Mathematical Analysis And Applications*. **395**, 336-354 (2012,11,1)
- [4] Denoyelle, Q., Duval, V., Peyré, G. & Soubies, E. The Sliding Frank–Wolfe Algorithm and Its Application to Super-Resolution Microscopy. *Inverse Problems*. **36**, 014001 (2020,1,1)
- [5] Golub, G. & Pereyra, V. Separable Nonlinear Least Squares: The Variable Projection Method and Its Applications. *Inverse Problems*. **19**, R1-R26 (2003,4,1)

Exact continuous relaxations of ℓ_0 -regularized criteria with non-quadratic data terms

- Luca Calatroni (Université Côte d’Azur, Laboratoire i3S,CNRS, France)
- **Mhamed Essafri** (Université de Toulouse, IRIT, CNRS, France)
- Emmanuel Soubies (Université de Toulouse, IRIT, CNRS, France)

Mots-clé : ℓ_0 regularization, non-convex optimization, non-quadratic data terms, exact relaxations.

Résumé : Sparse optimization approaches have attracted extensive attention due to their practical applications in various fields such as statistics, computer vision, signal/image processing, and machine learning. A typical formulation is given by

$$\hat{\mathbf{x}} \in \left\{ \arg \min_{\mathbf{x} \in \mathbb{R}^N} G_0(\mathbf{x}) := F_{\mathbf{y}}(\mathbf{A}\mathbf{x}) + \lambda \|\mathbf{x}\|_0 \right\}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ (with $M \ll N$) is a linear operator, $\mathbf{y} \in \mathbb{R}^M$ denotes the observed data, and the term $\|\cdot\|_0$ is the ℓ_0 pseudo-norm on \mathbb{R}^N that counts the number of non-zero components of its arguments. The function $F_{\mathbf{y}} : \mathbb{R}^M \rightarrow \mathbb{R}_{\geq 0}$ is a measure of fit between $\mathbf{A}\mathbf{x}$ and \mathbf{y} , and $\lambda > 0$ controls the trade-off between data fidelity and sparsity. Due to the nature of the ℓ_0 function, problem (1) is non-continuous, non-convex, and belongs to the class of NP-hard problems [1].

In this talk, we will introduce the ℓ_0 Bregman relaxation for (1), which is defined as a non-convex but continuous function providing a good approximation to the ℓ_0 term. First, we will present results on the existence of global minimizers. Then, we will show that replacing the ℓ_0 term in (1) with the ℓ_0 Bregman relaxation leads to an exact relaxation, denoted \tilde{G} , of G_0 . This generalizes existing work [2, 3] dedicated to the least squares case.

The relaxation \tilde{G} is said to be exact as it “reduces” the non-convexity of (1) (e.g., fewer local minimizers, wider basins of attraction) while preserving its global solution(s). Finally, a tailored Bregman proximal gradient algorithm is used for the numerical solution of the relaxed problem. Several numerical results for problems involving widely-used non-quadratic data terms (Kullback-Leibler divergence, logistic regression) will be shown.

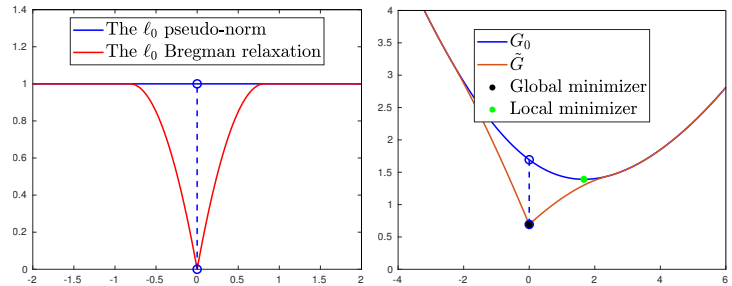


Figure 1: Illustration of the ℓ_0 Bregman relaxation for 1D logistic sparse regression.

Références

- [1] T. T. Nguyen, C. Soussen, J. Idier, and E-H. Djermoune. NP-hardness of ℓ_0 minimization problems: revision and extension to the non-negative setting. *Proceedings of SAMPTA, Bordeaux*, 2019.
- [2] Marcus Carlsson. On Convex Envelopes and Regularization of Non-convex Functionals Without Moving Global Minima. *Journal of Optimization Theory and Applications*, 183(1), 66–84, 2019.
- [3] E. Soubies, L. Blanc-Féraud, and G. Aubert. A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3), 2015.

Γ convergence of a N -player game to a Lagrangian Mean Field Game with pair-wise interaction

- João Miguel Machado (Université Paris-Dauphine, Inria Mokaplan)
- Guilherme Mazanti (Centrale Supelèc, Inria Disco)
- Laurent Pfeiffer (Centrale Supelèc, Inria Disco)

Mots-clé : Lagrangian MFGs, Γ convergence

Résumé : Let X and Y be Polish spaces and a continuous map $\pi: X \rightarrow Y$. To each agent is associated a $y \in Y$, following a distribution $m_0 \in \mathcal{P}(Y)$. The goal of an agent associated with $y \in Y$ is to

$$\underset{x \in X, \pi(x)=y}{\text{minimize}} \quad F(x, Q) = L(x) + \int_X H(x, \tilde{x}) dQ(\tilde{x})$$

where L is interpreted as an individual cost, H is a pairwise interaction and $Q \in \mathcal{P}(X)$ denotes the distribution of the choices of all agents. We say that $Q \in \mathcal{P}(X)$ is an *equilibrium* if $\pi_{\#}Q = m_0$, $\int_X F(x, Q) dQ(x) < +\infty$, and Q -almost every $x \in X$ minimizes $F(\cdot, Q)$, for $y = \pi(x)$.

It was shown in [2] that a particular case of this game has a potential structure; that is, its equilibria are minimizers of an energy. The arguments generalize to this abstract problem and Q is an equilibrium if and only if it

$$\underset{Q \in \mathcal{P}_{m_0}(X)}{\text{minimize}} \mathcal{J}(Q) \stackrel{\text{def.}}{=} \int_X L dQ + \int_{X \times X} H dQ \otimes Q \text{ where } \mathcal{P}_{m_0}(X) := \{Q \in \mathcal{P}(X) : \pi_{\#}Q = m_0\}.$$

This proposed abstract formulation unifies many problems already known in the literature. When $X = Y$, $\pi = \text{id}$ and $H(x, \tilde{x}) = g(x - \tilde{x})$, where g is the Green function of the Laplacian, it becomes the capacitor problem [3]. By taking $Y = \Omega$ a compact subset of \mathbb{R}^d and $X = C([0, 1]; \Omega)$ and $\pi = e_0$, the evaluation operator at time 0, L is the kinetic energy of a given curve and H is the Cucker-Smale interaction kernel to promote consensus in the velocity (as in a flock of birds that align their velocities) we recover the MFG proposed by Santambrogio and Shim [1].

We shall then discuss how a suitable family of N -players game with pair-wise interaction also enjoys a potential structure, so that Nash equilibria can be found by the minimization of a sequence functionals \mathcal{J}_N which depend on an empirical measure of the distribution m_0 . We then show that these functionals Γ converge to \mathcal{J} , and as a result we show that there is a sequence of Nash equilibria of the N -player games that converge to an equilibrium of the Lagrangian MFG.

Références

- [1] F. Santambrogio and W. Shim. A Cucker–Smale Inspired Deterministic Mean Field Game with Velocity Interactions. *SIAM Journal on Control and Optimization*, 59:4155–4187, 2021.
- [2] S. Arjmand. Mean field games with free final time. *PhD thesis, École Polytechnique*, 2022.
- [3] O. Frostman. Potentiel d'équilibre et capacité des ensembles. *PhD thesis, Gleeurp*, 1935.

The interplay between geometry and convergence in Bregman proximal methods

- **Waïss Azizian** (LJK, Univ. Grenoble Alpes)
- Franck Iutzeler (Univ. Grenoble Alpes)
- Jerome Malick (CNRS)
- Panyotis Mertikopoulos (CNRS)

Résumé : Bregman proximal methods [1, 2] are first-order (constrained) optimization algorithms whose non-Euclidean updates come from a distance-like functional, the Bregman divergence. In this communication, we will examine the last-iterate convergence rate of these methods — from mirror descent to mirror-prox and its optimistic variants — as a function of the local geometry induced by the Bregman divergence [4]. In the setting of variational inequalities, we will show that the local convergence rate depends sharply on the Legendre exponent [3], which measures the growth rate of the underlying Bregman function (Euclidean, entropic, or other) near a solution. For linearly-constrained problems, we will take a closer look at the convergence rate across the constraints that are active at the solution. In this case, the rate additionally depends on the position of the vector field relative to these constraints (see Figure 1). Finally, we will discuss versions of these results in the stochastic setting and their implications in the design of step-size schedules.

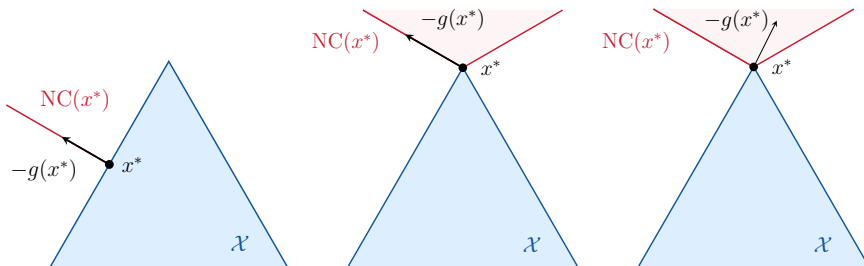


Figure 1: Different configurations give different convergence rates. In a linearly constrained problem: a non-extreme solution where the vector field g is sharp (left), an extreme solution where g is not sharp (center), and a sharp solution (right).

Références

- [1] A. S. Nemirovski and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.
- [2] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 1993.
- [3] W. Azizian, F. Iutzeler, J. Malick, and P. Mertikopoulos. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. *COLT*, 2021.
- [4] W. Azizian, F. Iutzeler, J. Malick, and P. Mertikopoulos. The rate of convergence of Bregman proximal methods: local geometry vs. regularity vs. sharpness. *In revision*, 2023.

Contrôle de volatilité locale par transport entropique

- Jean-David Benamou (INRIA Paris)
- **Guillaume Chazareix** (BNP Paribas, INRIA Paris)
- Marc Hoffmann (Université Paris-Dauphine)
- Grégoire Loeper (BNP Paribas)
- François-Xavier Vialard (Université Gustave Eiffel)

Mots-clé : Transport Optimal, Finance, Transport Entropique, Transport Martingale

Résumé : Le transport optimal martingale a trouvé de nombreuses applications dans divers contextes financiers, en particulier pour la calibration des processus stochastiques. En finance, on s'intéresse aux prix de produits financiers et de produits dérivés, qui sont eux-mêmes des produits financiers dont le prix est fonction du prix d'un ou plusieurs autres produits financiers. Leur prix est incertain, mais la distribution de ces prix est modélisée par une diffusion, par exemple une diffusion à coefficient constant dans le cas du célèbre modèle de Black-Scholes. Le problème de calibration correspond à trouver les paramètres d'un tel modèle en fonction de données mesurables sur le marché : le prix d'un produit financier à un instant initial et le prix de produits dérivés ayant maturité à des instants ultérieurs. Les produits dérivés sélectionnés seront généralement des produits courants, comme des options. Les modèles considérés peuvent avoir une volatilité locale, donc dépendante du temps et de l'espace. Cette volatilité peut ensuite être utilisée pour calculer le prix d'autres produits dérivés plus complexes. Les solutions numériques pour ce problème impliquent de résoudre un problème variationnel reposant sur des équations aux dérivées partielles non linéaires, tel que décrit dans [2]. Ces approches sont cependant limitées par la complexité de la résolution numérique de ces équations. Dans nos travaux, nous proposons une approche de discrétisation pour le problème continu et montrons que la solution du problème de transport optimal entropique martingale multi-marginales discret ainsi obtenu converge vers la solution pour le choix d'une fonction coût particulière du problème présenté dans [2]. Cette relaxation permet l'utilisation d'algorithmes analogues à ceux employés dans le transport optimal entropique classique, tel que présenté dans [1]. En outre, nous décrivons une méthode pour la mise en œuvre de cet algorithme sur une plate-forme GPU, améliorant ainsi la rapidité des calculs. Nous présentons des résultats numériques pour des exemples de transport martingale multi-marginale dans un contexte abstrait, ainsi que pour des exemples dans le cas particulier de la calibration de modèles de volatilité stochastique locale.

Références

- [1] Benamou, J.-D., Carlier, G., Di Marino, S., Nenna, L. An entropy minimization approach to second-order variational mean-field games. *Math. Models Methods Appl. Sci.*, 29, 1553-1583 (2019)
- [2] Guo, I., Loeper, G., Wang, S. Calibration of local-stochastic volatility models by optimal transport. *Mathematical Finance*, 32(1), 46-77 (2022)

Extrapolation dans l'espace de Wasserstein

- Thomas Gallouët (Inria Paris)
- Andrea Natale (Inria Lille)
- **Gabriele Todeschi** (Université Gustave Eiffel)

Mots-clé : Transport optimal, extrapolation, régularisation entropique

Résumé : Dans l'espace de Wasserstein $(\mathcal{P}(\Omega), W_2)$, c'est-à-dire l'espace des mesures de probabilité doté de la distance de Wasserstein W_2 , étant donné deux mesures de probabilité $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ ($\Omega \subseteq \mathbb{R}^d$, convexe), il existe toujours au moins une géodésique (à vitesse constante) $\mu : [0, 1] \rightarrow \mathcal{P}(\Omega)$, $\mu(0) = \mu_0$, $\mu(1) = \mu_1$. Cependant, ces courbes ne peuvent pas toujours être prolongées jusqu'à un temps $T > 1$ tout en restant une géodésique sur l'intervalle entier $[0, T]$. Dans [2], nous avons discuté de ce problème afin de construire un opérateur d'extrapolation à utiliser dans la discrétisation d'ordre élevé des flots de gradient de Wasserstein. Nous avons notamment proposé de définir la courbe extrapolée $\mu : [1, T] \rightarrow \mathcal{P}(\Omega)$ comme suit :

$$\mu(t) = \operatorname{argmin}_{\rho \in \mathcal{P}(\Omega)} tW_2^2(\rho, \mu_1) - (t-1)W_2^2(\rho, \mu_0), \quad t \in [1, T]. \quad (1)$$

Il s'agit d'une extension naturelle de la notion d'extrapolation euclidienne au cadre de Wasserstein. Pour $t \leq 1$, (1) est un problème d'optimisation convexe et l'on récupère une géodésique de μ_0 à μ_1 . Pour $t > 1$, le problème est encore bien posé, une solution existe toujours et est unique, mais ce n'est pas un problème d'optimisation convexe. Remarquablement, on peut montrer qu'il admet une formulation duale qui est convexe (de façon similaire à ce qui a été montré à l'origine dans [1]). Nous tirons parti de cette structure convexe pour concevoir une approche efficace et peu coûteuse en calcul pour le résoudre, basée sur la technique de régularisation entropique et une reformulation non triviale du problème, et calculer l'extrapolation $\mu(t)$.

Références

- [1] Guillaume Carlier. Remarks on Toland's duality, convexity constraint and optimal transport. 2008.
- [2] Thomas Gallouët, Andrea Natale, and Gabriele Todeschi. From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows. *arXiv preprint arXiv:2209.14622*, 2023.

The Mirror Descent Algorithm in Generalized Convexity

- **Adrien Le Franc** (LAAS CNRS)
- Jean-Philippe Chancelier (Cermics ENPC)
- Michel De Lara (Cermics ENPC)

Mots-clé : Generalized Convexity, Sparse Optimization

Résumé : In this talk, we present algorithmic perspectives based on mathematical tools of generalized Fenchel-Moreau conjugacies induced by *one-sided linear* (OSL) couplings.

We start with a generalization of the notion of Bregman divergence with OSL couplings, a first step before we also extend the mirror descent algorithm with OSL couplings [3]. This contribution is an original new step in the algorithmic toolbox of general convexity, dominated so far by the cutting angle algorithm [1].

We also comment on the possibility to apply our generalized mirror descent algorithm to sparse optimization problems that display Capra-convex properties [2]. In particular, we consider problems displaying the ℓ_0 sparse pseudonorm as an objective function. Numerical perspectives are encouraged by the knowledge of a closed-form expression for the Capra-subdifferential of ℓ_0 [4], which is requested for both the cutting angle and the mirror descent algorithm.

Références

- [1] A. Rubinov. Abstract convexity and global optimization. Nonconvex Optimization and its Applications, vol.44. Kluwer Academic Publishers, Dordrecht, 2000.
- [2] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the ℓ_0 pseudonorm, 2020. Optimization, 2022, vol. 71, no 2, p. 355-386.
- [3] A. Le Franc. Subdifferentiability in convex and stochastic optimization applied to renewable power systems. PhD Thesis, École des Ponts ParisTech, 2021.
- [4] A. Le Franc, J.-P. Chancelier and M. De Lara. The Capra-subdifferential of the ℓ_0 pseudonorm. Optimization, 2022.

Nonsmoothness in Wasserstein distributionally robust models

- **Tam Le** (LJK, Université Grenoble Alpes)

Mots-clé : machine learning, optimal transport, Wasserstein distance, distributionally robust optimization

Résumé : The estimation of machine learning models is often prone to overfitting and empirical risk minimization is sensitive to distribution shifts between training and deployment. Wasserstein distributionally robust optimization [3, 2] has emerged as an attractive solution to address this issue, leveraging optimal transport to enhance robustness and generalization capabilities of machine learning models. However, the statistical properties of the Wasserstein distance suffer from the curse of dimensionality, and fail to provide tight generalization guarantees. A recent work [1] highlights that better generalization guarantees hold under a uniform lower bound on the dual variable. We follow this approach and despite the nonsmooth nature of the problem, we establish, using a graphical construction (Figure 1), the existence of a lower bound to the dual solution, for a reasonable range of Wasserstein radius. This result enables us to derive generalization guarantees in most practical settings, encompassing general transport costs and non-differentiable objective functions prevalent in deep learning.

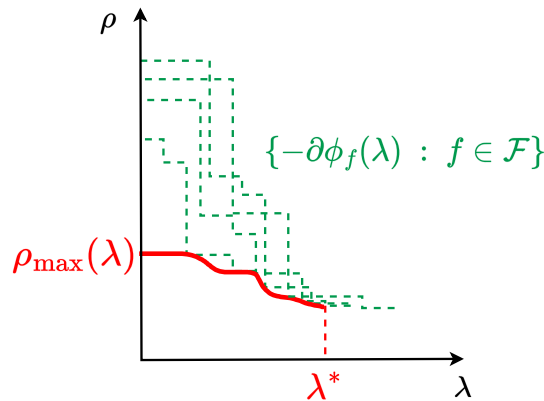


Figure 1: Graphical construction of the lower bound on the dual variable

Références

- [1] W. Azizian, F. Iutzeler & J. Malick, Exact Generalization Guarantees for (Regularized) Wasserstein Distributionally Robust Models. *NeurIPS* (2023)
- [2] J. Blanchet & K. Murthy, Quantifying distributional model risk via optimal transport. *Mathematics Of Operations Research*. **44**, 565-600 (2019)
- [3] P. Mohajerin Esfahani & D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* **171**, 115–166 (2018). <https://doi.org/10.1007/s10107-017-1172-1>

Optimisation du cycle limite d'un cerf-volant

- **Antonin Bavoil** (CNRS, Inria, Université Côte d'Azur, Nice, France)
- Jean-Baptiste Caillau (Université Côte d'Azur, CNRS, Inria, LJAD, Nice, France)
- Lamberto Dell'Ecce (Inria, Valbonne, France)
- Alain Nême (ENSTA Bretagne, Brest, France)
- Jean-Baptiste Leroux (ENSTA Bretagne, Brest, France)

Mots-clé : airborne wind energy, cycle limite

Résumé :

Les éoliennes sont actuellement le seul moyen de convertir l'énergie des vents en électricité, mais elles sont loin d'être parfaites : elles nécessitent une quantité importante de matériaux et le transport des composants sur le site de construction devient de plus en plus difficile en raison de leur taille croissante. Les voiles de cerf-volant résolvent ces deux problèmes car la production d'électricité nécessiterait jusqu'à 90 % de matériaux en moins pour la même quantité d'énergie produite, et l'absence de longs composants rigides rend le transport suffisamment facile pour déplacer la production d'énergie en fonction des besoins. La taille des éoliennes est limitée par le transport de leurs composants, alors qu'une voile de cerf-volant n'a besoin que de lignes plus longues pour voler plus haut, ce qui permet de bénéficier de vents en altitude plus forts et plus réguliers. Cependant, le cerf-volant a besoin d'une commande qui le maintienne en l'air et maximise la production d'énergie. L'objectif actuel du projet KEEP (Kite Electric Energy Production), piloté par l'ENSTA Bretagne en collaboration avec Université Côte d'Azur, est de simuler le système afin d'obtenir une première estimation de la production d'électricité.

Le modèle utilisé est une équation différentielle ordinaire du second ordre à deux variables, la commande étant émulée par une force fictive de sorte à ce que le cerf-volant se déplace sur une variété prédéterminée — en projection, une courbe en forme de huit. La solution de ce problème converge vers un cycle limite, et on s'intéresse à l'énergie produite durant ce cycle. On s'intéresse à la maximisation de l'énergie sur ce cycle en fonction des paramètres mécaniques du cerf-volant, sous la contrainte non-linéaire que la tension des lignes reste en dessous d'un certain seuil [1]. On propose également une modélisation plus fine faisant intervenir une commande par l'angle d'incidence.

Références

- [1] K. Desenclos, A. Nême, J.B. Leroux, C. Jochum. A novel composite modelling approach for woven fabric structures applied to leading edge inflatable kites. *Mechanics of Composite Materials*, Volume 58, issue 6, 2023

Long-time asymptotics of the Sliced Wasserstein Gradient Flow

- Filippo Santambrogio (Univ. C. Bernard 1, Lyon)
- **Giacomo Cozzi** (Univ. degli Studi di Padova)

Mots-clé : Optimal Transport, Wasserstein Spaces, Gradient Flows

Résumé : We discuss the properties of the gradient flow associated with the Sliced Wasserstein distance (an useful-for-applications variant of the usual Wasserstein distance) with respect to a fixed target measure, in the space of probability measures over \mathbb{R}^d , endowed with the 2-Wasserstein metric. We focus on convergence issues for the flow and, starting by some ideas contained in [2], we show that the limit of the map describing the Lagrangian characteristics is not the optimal transport map from the starting measure to the target.

Références

- [1] F. Santambrogio. Optimal Transport for Applied Mathematicians *Birkäuser, NY*, 94 (2015).
- [2] H. Lavenant & F. Santambrogio. The flow map of the Fokker–Planck equation does not provide optimal transport *Applied Mathematics Letters*. 133 pp. 108225 (2022)
- [3] N. Bonnotte. Unidimensional and evolution methods for optimal transportation. (Université Paris Sud-Paris XI; Scuola normale superiore (Pise, Italie),2013)

Entropic Optimal Transport Solutions of the Semigeostrophic Equations

- Jean-David Benamou (Inria Paris)
- Colin Cotter (Imperial College)
- **Hugo Malamut** (Ceremade)

Mots-clé : équations semi-géostrophiques, transport optimal entropique, Sinkhorn algorithm

Résumé : Dans le domaine de la prévision atmosphérique, l'hypothèse du 'slow manifold' postule l'existence d'un modèle valide à grande échelle spatiale et temporelle, qui échapperait au chaos, au sens où il serait peu sensible à des variations de petite échelle. C'est dans ce contexte que les équations semi-géostrophiques prennent toute leur importance. En effet, elles correspondent au modèle de slow manifold le plus satisfaisant, solvable en temps long et valide asymptotiquement à grande échelle, du moins loin de l'équateur ([1]).

Or, sous une hypothèse de stabilité, l'existence de solutions, tant du point de vue théorique que numérique, est liée au transport optimal. En effet, un changement de variable permet d'interpréter le gradient de pression comme une application de transport optimal entre la densité du fluide en coordonnées physiques et géostrophiques. Sous cette forme, les équations prennent alors la forme d'un système hamiltonien élémentaire dans l'espace de Wasserstein.

D'un point de vue numérique, le transport optimal computationnel s'est développé rapidement durant les deux dernières décennies. Parmi les méthodes les plus populaires, on trouve l'algorithme de Sinkhorn, qui permet de calculer les solutions d'une régularisation entropique du problème de transport optimal. Dans cet exposé, je présenterai les résultats de [2], où nous montrons sur un exemple classique (Eady slice model) que l'algorithme de Sinkhorn semble prometteur pour calculer numériquement les solutions des équations semi-géostrophiques. Si le temps le permet, je discuterai également des questions d'existence et d'unicité du système semi-géostrophique entropisé, qui correspondent à un travail en cours avec Guillaume Carlier.

Références

- [1] M. J. P. Cullen. A mathematical theory of large-scale atmosphere/ocean flow. *World Scientific*, 2006.
- [2] J. D. Benamou, C. Cotter, and H. Malamut. Entropic Optimal Transport Solutions of the Semigeostrophic Equations. Working paper or preprint, January 2023.

Random subspaces and expected decrease in derivative-free optimization

- Warren Hare (University of British Columbia, Canada)
- Lindon Roberts (University of Sydney, Australia)
- **Clément W. Royer** (Université Paris Dauphine-PSL, France)

Mots-clé : derivative-free optimization, random subspaces, decrease guarantees

Résumé :

Derivative-free algorithms seek the minimum of a given function based only on function values queried at appropriate points. Although these methods are widely used in practice, their performance is known to worsen as the problem dimension increases. Recent advances in developing randomized derivative-free techniques have tackled this issue by working in low-dimensional subspaces that are drawn at random in an iterative fashion. The connection between the dimension of these random subspaces and the algorithmic guarantees has yet to be fully understood.

In this talk, we describe several strategies to select random subspaces within a derivative-free algorithm. We focus on direct search for simplicity, and explain how probabilistic convergence rates can be obtained for such a method. Experiments for these methods show a clear advantage for using one-dimensional random subspaces. We investigate this behavior through a novel, expected decrease analysis that highlights a connection between subspace dimension and per-iteration decrease guarantees.

Références

- [1] W. Hare, L. Roberts and C. W. Royer.
Expected decrease for derivative-free algorithms using random subspaces.
Technical report arXiv:2308.04734.
- [2] L. Roberts and C. W. Royer.
Direct search based on probabilistic descent in reduced spaces.
SIAM Journal on Optimization,
33(4):3057-3082, 2023.

Sous-régularité métrique et prox-régularité

- Florent Nacry (Université Perpignan)
- Vo Anh Thuong Nguyen (Université Perpignan)
- Juliette Venel (Université Polytechnique Hauts-de-France)

Mots-clé : Analyse variationnelle, régularité métrique, prox-régularité, équations généralisées

Résumé : Nous débutons cet exposé en rappelant l'importance de la régularité métrique en optimisation. Disons qu'une multi-application $M : X \rightrightarrows Y$ est *métriquement régulière* en un point (\bar{x}, \bar{y}) de son graphe lorsqu'il existe un réel $\gamma \geq 0$ et un voisinage V de (\bar{x}, \bar{y}) tels que

$$d(x, M^{-1}(y)) \leq \gamma d(y, M(x)) \quad \text{pour tout } (x, y) \in V,$$

où la notation $d(\cdot, S)$ désigne la distance à un ensemble S . Nous montrons que cette inégalité avec $y = \bar{y}$ et x au voisinage de \bar{x} (qui constitue la notion de *sous-régularité métrique*) permet d'obtenir la normale $\omega(\cdot)$ -régularité (dont un cas particulier bien connu est la prox-régularité hilbertienne) d'équations généralisées $\{x \in X : f(x) \in M(x)\}$. Rappelons qu'une partie S d'un Hilbert H est *prox-régulière* de constante r lorsque la projection métrique sur S est bien définie (au sens de l'existence et de l'unicité) en chaque point du sous-niveau $\{d_S < r\}$. Ceci est équivalent à un certain comportement *hypomonotone* du cône normal $N(S; \cdot)$ à l'ensemble S :

$$\langle \zeta, x' - x \rangle \leq \frac{\|\zeta\|}{2r} \|x' - x\|^2 \quad \text{pour tout } x, x' \in S, \text{ pour tout } \zeta \in N(S; x).$$

Enfin, nous développons des résultats de type Robinson-Ursescu assurant la sous-régularité métrique sous une condition d'ouverture. Plus précisément, lorsque le graphe de M est prox-régulier de constante r et qu'il existe deux constantes $\alpha, \beta > 0$ tels que

$$B(\bar{y}, \beta) \subset M(B[\bar{x}, \alpha]) \quad \text{et} \quad \beta > \frac{\alpha^2 + \beta^2}{2r}$$

nous obtenons que la multi-application M est sous-métriquement régulière au sens donné ci-dessus. Ces résultats sont notamment appliqués pour obtenir de nouvelles conditions assurant la prox-régularité de l'intersection d'ensembles prox-réguliers ou encore la prox-régularité d'ensembles de contraintes $\{f_1 \leq 0, \dots, f_m \leq 0, f_{m+1} = 0, \dots, f_{m+n} = 0\}$.

Références

- [1] S. Adly, F. Nacry, L. Thibault, Prox-regularity approach to generalized equations and image projection. *ESAIM: Control, Optimisation and Calculus of Variations*, 24:677–708, 2018.
- [2] F. Nacry, V.A.T. Nguyen, J. Venel, Metric subregularity and $\omega(\cdot)$ -normal regularity properties. *Submitted to JOTA*.
- [3] L. Thibault Unilateral Variational Analysis in Banach spaces. Part 1 & Part 2. *World Scientific*, 2023.
- [4] J. Venel A numerical scheme for a class of sweeping processes. *Numer. Math.* 118:367–400, 2011.
- [5] X.Y. Zheng, K.F. Ng Metric subregularity for proximal generalized equations in Hilbert spaces. *Nonlinear Anal.* 75 75:1686–1699, 2012.

Méthodes d'ordre élevées sans évaluation de la fonction objective

- **Sadok Jerad** (Université de Toulouse)
- Serge Gratton (Université de Toulouse)
- Philippe L. Toint (Université de Naxys)

Mots-clé : Optimisation Nonconvexe, Méthodes d'Ordre Élevée, Optimisation sans Fonction Objective, Optimisation Stochastique

Résumé : Le développement des outils d'intelligence artificielle s'est accompagné de développement de nouvelles méthodes d'optimisation. Ces nouveaux algorithmes du premier ordre comme Adam [3], n'utilisent pas la valeur fonction objective pour la mise à jour de leur paramètres contrairement aux méthodes plus standards comme la recherche linéaire ou les régions de confiance. Se basant sur le succès de ces méthodes du premier ordre qui n'évaluent pas la fonction et que des études théoriques indiquent que les algorithmes bruités sont très sensibles à la valeur de la fonction objective [2], nous avons proposé en le premier algorithme à tenseurs exacte qui n'évaluent pas le fonction objective et gardent la complexité des méthodes standards [1]. Des expériences numériques initiales semblent confirmer la justesse notre approche, en particulier pour les problèmes bruités.

Références

- [1] Serge Gratton, Sadok Jerad, and Philippe L. Toint. Convergence properties of an objective-function-free optimization regularization algorithm, including an $\mathcal{O}(\epsilon^{-3/2})$ complexity bound. *SIAM Journal on Optimization*, 33(3):1621–1646, 2023.
- [2] S. Bellavia, G. Gurioli, B. Morini, and Philippe L. Toint. Adaptive regularization for nonconvex optimization using inexact function values and randomly perturbed derivatives. *Journal of Complexity*, February 2022.
- [3] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings in the International Conference on Learning Representations (ICLR)*, 2015.

Unrolled algorithms in the SPD cone and application to sparse graph learning

- **Can Pouliquen** (Univ Lyon, Ens Lyon, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France.)
- Paulo Gonçalves (Univ Lyon, Ens Lyon, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France.)
- Titouan Vayer (Univ Lyon, Ens Lyon, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France.)
- Mathurin Massias (Univ Lyon, Ens Lyon, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France.)

Mots-clé : Unrolled algorithms, Positive-definiteness, Sparsity, Graphical Lasso

Résumé : Inferring sparse graphical representations of data is of prime interest, and various methods seek to estimate a sparse positive-definite (PD) matrix that characterizes the graph connections. While convex optimization algorithms can offer these estimates [1], recent progress in deep unrolled algorithms shows promising potential. However, when it comes to sparse PD matrix estimation, maintaining the sparsity, symmetry, and positive-definiteness of every layer outputs in these algorithms remains challenging [2]. In this paper, we introduce a generic framework for unrolling these optimization algorithms, which effectively maintains these properties across the training process. We illustrate how our framework can be applied to address the well-known challenges associated with unrolling different graph inference algorithms, as well as how it can be extended to enhance their flexibility in a Plug-and-Play [3] fashion. Ultimately, we showcase the versatility of our framework through our experiments.

Références

- [1] J. Friedman, T. Hastie and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 3:432–441, 2008.
- [2] H. Shrivastava, X. Chen, B. Chen, G. Lan, S. Aluru, H. Liu and L. Song. GLAD: Learning sparse graph recovery. *International Conference on Learning Representations (ICLR)*, 2020.
- [3] S. V. Venkatakrisnan, C. A. Bouman B. Wohlberg. Plug-and-play priors for model based reconstruction. *2013 IEEE global conference on signal and information processing*, 945–948, 2013.

A Comparison Principle for Semi-Linear Hamilton-Jacobi Equations in the Wasserstein Space

- **Samuel Daudin** (Université Côte d'Azur)
- Benjamin Seeger (University of Texas at Austin)

Mots-clé : Mean-Field Control, Hamilton-Jacobi equation, Wasserstein Space

Résumé : I will present a comparison principle for viscosity solutions of semilinear Hamilton-Jacobi equations in the space of probability measures. Such equations arise, for instance, as the dynamic programming equations for the optimal control of McKean-Vlasov SDEs. Our method involves leveraging differentiability properties of the 2-Wasserstein distance in the doubling of variables argument, which is done by introducing a further entropy penalization that ensures that the relevant optima are achieved at positive, Lipschitz continuous densities with finite Fisher information. This allows to prove uniqueness and stability of viscosity solutions in the class of bounded Lipschitz continuous (with respect to the 1-Wasserstein distance) functions. The result does not appeal to a mean field control formulation of the equation, and, as such, applies to equations with non-convex Hamiltonians and measure-dependent volatility. For convex Hamiltonians that derive from a potential, we prove that the value function associated with a suitable mean-field optimal control problem with nondegenerate idiosyncratic noise is indeed the unique viscosity solution.

Based on the joint work [1] with Ben Seeger.

Références

- [1] S. Daudin and B. Seeger A Comparison Principle for Semilinear Hamilton Jacobi Equations in the Wasserstein Space *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

Exact continuous relaxations of ℓ_0 -regularized criteria with non-quadratic data terms

- **Mhamed Essafri** (Université de Toulouse, IRIT, CNRS, France)
- Luca Calatroni (Université Côte d’Azur, Laboratoire i3S, CNRS, France)
- Emmanuel Soubies (Université de Toulouse, IRIT, CNRS, France)

Mots-clé : ℓ_0 regularization, non-convex optimization, non-quadratic data terms, exact relaxations.

Résumé : Sparse optimization approaches have attracted extensive attention due to their practical applications in various fields such as statistics, computer vision, signal/image processing, and machine learning. A typical formulation is given by

$$\hat{\mathbf{x}} \in \left\{ \arg \min_{\mathbf{x} \in \mathbb{R}^N} G_0(\mathbf{x}) := F_{\mathbf{y}}(\mathbf{A}\mathbf{x}) + \lambda \|\mathbf{x}\|_0 \right\}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ (with $M \ll N$) is a linear operator, $\mathbf{y} \in \mathbb{R}^M$ denotes the observed data, and the term $\|\cdot\|_0$ is the ℓ_0 pseudo-norm on \mathbb{R}^N that counts the number of non-zero components of its arguments. The function $F_{\mathbf{y}} : \mathbb{R}^M \rightarrow \mathbb{R}_{\geq 0}$ is a measure of fit between $\mathbf{A}\mathbf{x}$ and \mathbf{y} , and $\lambda > 0$ controls the trade-off between data fidelity and sparsity. Due to the nature of the ℓ_0 function, problem (1) is non-continuous, non-convex, and belongs to the class of NP-hard problems [1].

In this talk, we will introduce the ℓ_0 Bregman relaxation for (1), which is defined as a non-convex but continuous function providing a good approximation to the ℓ_0 term. First, we will present results on the existence of global minimizers. Then, we will show that replacing the ℓ_0 term in (1) with the ℓ_0 Bregman relaxation leads to an exact relaxation, denoted \tilde{G} , of G_0 . This generalizes existing work [2, 3] dedicated to the least squares case.

The relaxation \tilde{G} is said to be exact as it “reduces” the non-convexity of (1) (e.g., fewer local minimizers, wider basins of attraction) while preserving its global solution(s). Finally, a tailored Bregman proximal gradient algorithm is used for the numerical solution of the relaxed problem. Several numerical results for problems involving widely-used non-quadratic data terms (Kullback-Leibler divergence, logistic regression) will be shown.

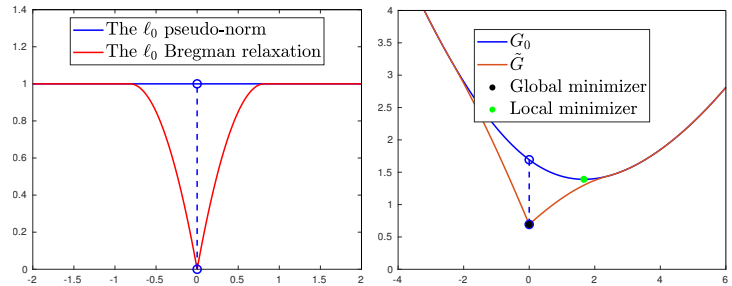


Figure 1: Illustration of the ℓ_0 Bregman relaxation for 1D logistic sparse regression.

Références

- [1] T. T. Nguyen, C. Soussen, J. Idier, and E-H. Djermoune. NP-hardness of ℓ_0 minimization problems: revision and extension to the non-negative setting. *Proceedings of SAMPTA, Bordeaux*, 2019.
- [2] Marcus Carlsson. On Convex Envelopes and Regularization of Non-convex Functionals Without Moving Global Minima. *Journal of Optimization Theory and Applications*, 183(1), 66–84, 2019.
- [3] E. Soubies, L. Blanc-Féraud, and G. Aubert. A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3), 2015.

Optimality conditions for parabolic stochastic optimal control problems with boundary controls

- **Piero Visconti** (LMI - INSA Rouen Normandie)

Mots-clé : Stochastic optimal control, Boundary controls, Necessary optimality conditions, Backward stochastic evolution equations

Résumé : Optimality conditions are provided for a class of control problems driven by a cylindrical Wiener process, which amount to a stochastic maximum principle in differential form. The control is considered to act on the drift and the volatility, both of which may be unbounded operators, which allows us to consider SPDEs with control and/or noise on the boundary. By the factorization method (developed for linear equations in [1]), a regularizing property is established for the state equation which is then employed to prove, by duality, a similar result for the backward time costate equation. The costate equation is understood in the sense of transposition (see [2]). Finally, the cost is shown to be Gateaux differentiable and its derivative is represented in terms of the costate, the optimality condition is deduced using results of set valued analysis (procedure developed in [3] to obtain pointwise conditions from variational conditions).

Références

- [1] G. Da Prato, S. Kwapień & J. Zabczyk. Regularity of solutions of linear stochastic equations in hilbert spaces. *Stochastics: An International Journal of Probability and Stochastic Processes*, 23(1):1–23, 1988.
- [2] Q. Lü, X. Zhang. Mathematical control theory for stochastic partial differential equations. *Probability Theory and Stochastic Modeling*, 101, 2021.
- [3] T. Wang, H. Zhang. Optimal control problems of forward-backward stochastic Volterra integral equations with closed control regions. *SIAM Journal on Control and Optimization*, 55(4):2574–2602, 2017.

Improving Sparse Penalties with Nonconvexity and Coefficient Clustering

- **Anne Gagneux** (Ens Lyon, Inria, LIP, Lyon, France)
- Mathurin Massias (Ens Lyon, Inria, LIP, Lyon, France)
- Emmanuel Soubies (CNRS, IRIT, Univ Toulouse, Toulouse, France)

Mots-clé : sparsity, nonconvexity, variable clustering, regularization

Résumé : In this work, we address the following question:

how can we enforce parsimony and structure in a regularized regression model ?

Specifically, we investigate strategies that promote not only sparsity but also clustering of correlated features. In fact, we aim to recover these groups *without prior knowledge about the clusters*. For such problems, the Sorted ℓ_1 norm (SLOPE [1], OWL [2]) is already widely used. Yet, SLOPE relies on the ℓ_1 penalty. It therefore shares a major drawback with the Lasso: it is biased and tends to shrink non-null coefficients.

In the same way nonconvex penalties have been introduced as unbiased ℓ_1 alternatives, we deal here with *sorted nonconvex penalties as unbiased SLOPE's counterparts*. Feng and Zhang [3] have already proposed an approach to address a family of sorted nonconvex penalties: they use a "majorization-minimization" method that generates a sequence of easier-to-solve convex surrogate problems.

Instead, we show that for a parameter η small enough in $\text{prox}_{\eta\Psi}$, we can exactly compute the prox of these penalties and apply proximal gradient algorithms to solve the penalized regression. We introduce a new formulation for the prox of sorted penalties which only involves the scalar prox of the penalty (for which we usually have an explicit expression) and averaged values of parts of the input vector and the regularization sequence.

We experimentally show that using sorted nonconvex penalties does improve both unbiasedness of the solution and automatic features grouping, which validates our initial hypothesis.

Références

- [1] Bogdan, Malgorzata, et al, SLOPE-adaptive variable selection via convex optimization, *The annals of applied statistics* 9.3 (2015): 1103.
- [2] Zeng, Xiangrong, and Mario AT Figueiredo, The Ordered Weighted ℓ_1 Norm: Atomic Formulation, Projections, and Algorithms, arXiv preprint arXiv:1409.4271 (2014).
- [3] Feng, Long, and Cun-Hui Zhang, Sorted concave penalized regression, *The Annals of Statistics* 47.6 (2019).

Adapting to game trees in zero-sum imperfect information games

- Côme Fiegel (CREST, ENSAE Paris)
- Pierre Ménard (ENS Lyon)
- Tadashi Kozuno (Omron Sinic X)
- Rémi Munos (Deepmind Paris)
- Vianney Perchet (CREST, ENSAE Paris)
- Michal Valko (Deepmind Paris)

Mots-clé : Sequential learning, Imperfect information

Résumé : Imperfect information games are games in which each player only partially observes the current game state. We study how to learn ϵ -optimal strategies in a zero-sum game through self-play with *trajectory feedback*. We give a problem-independent lower bound $\tilde{O}(H(A_{\mathcal{X}} + B_{\mathcal{Y}})/\epsilon^2)$ on the required number of realizations to learn these strategies with high probability, where H is the length of the game, $A_{\mathcal{X}}$ and $B_{\mathcal{Y}}$ are the total number of actions for the two players. We also propose two Follow the Regularized leader (FTRL) algorithms for this setting: **Balanced FTRL** which matches this lower bound, but requires the knowledge of the information set structure beforehand to define the regularization; and **Adaptive FTRL** which needs $\tilde{O}(H^2(A_{\mathcal{X}} + B_{\mathcal{Y}})/\epsilon^2)$ realizations without this requirement by progressively adapting the regularization to the observations.

Références

- [1] Côme Fiegel, Pierre Ménard, Tadashi Kozuno, Rémi Munos, Vianney Perchet, Michal Valko. Adapting to game trees in zero-sum imperfect information games. *International Conference on Machine Learning 2023*

A General Theory for Exact Sparse Representation Recovery in Convex Optimization

- **Leonardo Del Grande** (University of Twente)
- Marcello Carioni (University of Twente)

Mots-clé : Sparsity, extreme points, metric space

Résumé : In this contribution we analyze the recovery of the sparse representation of data in general infinite-dimensional optimization problems regularized by convex functionals.

By assuming a suitable non-degeneracy condition on the problem we establish that, for small regularization parameters and noise levels, the minimizer is unique and is uniquely represented as a linear combination of n extreme points of the ball of the regularizer. Such non-degeneracy condition extends the classical non-degeneracy source condition (NDSC) for total variation regularized inverse problems introduced by Duval and Peyré [1]. More precisely, it is connected to the behaviour of the solution of the dual problem when evaluated on the set of extreme points of the ball of the regularizer, seen as a metric space. This justifies the name Metric Non-Degenerate Source Condition (MNDSC).

Finally, we obtain explicit formulations of the MNDSC, which lead us to specific results of sparse recovery for three problems of interest:

- Total variation regularized deconvolution problems, where we show that the classical NDSC implies our MNDSC;
- 1-dimensional BV functions regularized with their BV-seminorm;
- Pairs of measures regularized with their mutual 1-Wasserstein distance.

A lot of other problems can be explored applying this result, particularly those considering optimal transport-type regularizers, as presented in our last example. Of particular interest for future works are dynamic problems regularized by the Benamou-Brenier energy, where extreme points have been already identified in [2]. This contribution is based on the preprint [3], joint work with Marcello Carioni.

Références

- [1] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2014.
- [2] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. On the extremal points of the ball of the Benamou–Brenier energy. *Bulletin of the London Mathematical Society*, 53(5):1436–1452, 2021.
- [3] L. Del Grande and M. Carioni. A General Theory for Exact Sparse Representation Recovery in Convex Optimization. *arXiv preprint arXiv:2311.08072*, 2023.

Optimisation géométrique de formes pour l'optique réflexive

- Samuel Amstutz (LMA, Avignon Université)
- Benjamin Aymard (Thales Alenia Space, Cannes)
- Audric Drogoul (Thales Alenia Space, Cannes)
- **Luca Gorini** (CMAP, École Polytechnique)

Mots-clé : optimisation de formes, optique géométrique, variable aléatoires, méthode de variation de frontière, réflexion spéculaire, problème adjoint

Résumé : Un télescope optique a pour objectif de former l'image d'un objet lointain comme la Terre ou une étoile dans un plan focal donné. Pour cela, la lumière collectée par l'instrument traverse divers miroirs et lentilles. Ces éléments permettent de re-focaliser l'image en sortie du système. La qualité de l'image ainsi reformée dépend fortement de la forme des miroirs et des lentilles. Nous nous limitons au cas des miroirs.

Jusqu'à présent, de nombreux télescopes optiques sont axisymétriques. De nombreux miroirs sont coniques et invariants par rotation autour d'un axe dit optique. Les coniques présentent des propriétés optiques connues et sont facilement usinables. Néanmoins des résultats récents [1] ont montré que s'affranchir de l'invariance par rotation et des coniques dans la forme des miroirs permet d'améliorer certaines limitations (obscuration, augmentation de luminosité, ...). Nous étudions le problème d'optimisation de la forme de ces miroirs.

Diverses approches pour traiter l'optique géométrique dans les miroirs existent en théorie comme en pratique [2]. Une approche simple consiste à modéliser la source lumineuse à l'aide de variables aléatoires décrivant la position et la direction des rayons. Une fois les rayons émis, ces derniers suivent un trajet déterministe constitué de trajectoires rectilignes et de réflexions spéculaires (en nombre fini). Une fonction coût J écrite dans le plan focal rend compte de la qualité de l'image. Ce coût dépend des points et directions d'impacts des rayons sur le plan focal, eux-mêmes influencés par la forme des miroirs. On cherche à minimiser J par rapport à la forme des miroirs. Pour cela, nous utilisons la méthode de variation de frontière d'Hadamard [3]. Nous adaptons les techniques de calcul de forme développées dans un contexte d'EDP à notre modèle stochastique. Nous présentons des résultats numériques qui permettent de retrouver les formes bien connus de l'optique pour 1 miroir (parabole, ellipse et hyperbole).

Références

- [1] A. Bauer and E. Schiesser and J. Rolland. Starting geometry creation and design method for freeform optics. *Nature Communications*, 2018.
- [2] O. Runborg. Computational high frequency wave propagation. *Cambridge University Press*, 2003.
- [3] A. Henrot and M. Pierre. Variation et optimisation de formes: Une analyse géométrique. *Springer Berlin Heidelberg*, 2005.

Maximum of Reachable Values of Discrete-Time Systems

- Assalé Adjé (Université de Perpignan Via Domitia, France) assale.adje@univ-perp.fr

Mots-clé : Discrete-time system, Sequences suprema, Lyapunov functions

Résumé : In this presentation, we are interested in finding a reachable value of a discrete-time dynamical system that maximizes a real valued function. This kind of maximization problems arises in the analyze of the robustness or the performance of the system; for inverse problems or even in formal verification of systems. The optimal value in this context can be viewed as the value which penalizes the most the system with respect to some criteria and is used as an extreme situation where the system enters.

Mathematically, the problem is formulated from an objective real-valued continuous function and a discrete-time dynamical system. This latter is characterized by, the dynamics which is a self map on a finite dimensional vector space and an initial conditions set which is a compact subset. The constraints set of this maximization problem is exactly the set of the reachable values of the system i.e. the set of terms of the recurrence sequences starting from an element in the initial conditions set. In this context, the optimal solution is composed of an initial condition and an integer which represents the rank for which the term is maximal. The compactness of the initial conditions set and adding some continuity assumptions constitute a comfortable framework for the optimal initial condition existence. The difficulty of the problem lies on the characterization of the optimal integer solution.

Then, in this presentation, we propose characterizations of the optimal integer solution based on homeomorphisms of geometric sequences in a total general setting. When the system is Lyapunov stable, such a homeomorphism can be defined from Lyapunov functions. Then computable upper-bounds of the optimal integer solution of our problem derive from the computation of Lyapunov functions. We illustrate our approach on discrete-time stable affine dynamical systems for which semidefinite programming can be employed to compute such upper-bounds [1]. We also aboard the problem of the computation of the minimal upper-bound and its formulation as a nonlinear semidefinite program.

Références

- [1] A. Adjé, Quadratic Maximization of Reachable Values of Stable Discrete-Time Affine Systems. *preprint*, <https://doi.org/10.48550/arXiv.1810.09778>, 2023.

Méthodes proximales multi-niveaux pour la restauration d'images

- **Guillaume Lauga** ([†]Univ Lyon, Inria, EnsL, UCBL, CNRS, LIP, UMR 5668, F-69342)
- Elisa Riccietti ([†])
- Nelly Pustelnik (Ens de Lyon, CNRS, Laboratoire de Physique, F-69342)
- Paulo Gonçalves ([†])

Mots-clé : méthodes proximales, optimisation multi-niveau, restauration d'images

Résumé : La résolution de problèmes d'optimisation en grande dimension est une tâche difficile, et de nombreuses méthodes ont été proposées pour pallier le coût important en temps de calcul. La piste explorée ici est l'exploitation de la structure de ces problèmes d'optimisation.

Dans ce travail, nous nous concentrons sur la structure multi-résolution au cœur des méthodes d'optimisation multi-niveaux. Ces approches tirent parti de la définition d'approximations grossières de la fonction objective pour rendre plus efficace sa minimisation. Dans cet exposé, nous présentons un algorithme proximal multi-niveaux IML FISTA - Inexact Multilevel FISTA - applicable à la résolution de problèmes d'optimisation dont la composante non-lisse n'a pas de formulation explicite pour son opérateur proximal [1].

La méthode proposée est ensuite étendue pour résoudre des problèmes liés à la restauration d'images hyperspectrales [2]. En mettant à profit la nature de ces images qui montrent de fortes corrélations dans le plan spectral, nous démontrons le potentiel d'IML FISTA à produire des images restaurées en un temps de calcul bien plus faible que les méthodes de l'état-de-l'art.

Références

- [1] G. Lauga, E. Riccietti, N. Pustelnik, P. Gonçalves. IML FISTA: A Multilevel Framework for Inexact and Inertial Forward-Backward. Application to Image Restoration. *Preprint*, 2023.
- [2] G. Lauga, E. Riccietti, N. Pustelnik, P. Gonçalves. Méthodes multi-niveaux pour la restauration d'images hyperspectrales. *Colloque GRETSI*, 2023.

Préconditionnement géométrique de la méthode indirecte de tir appliquée au véhicule hybride

- Olivier Cots (IRIT/INP-ENSEEIH)
- **Rémy Dutto** (Vitesco Technologies, IRIT/INP-ENSEEIH, IMT)
- Sophie Jan (IMT)
- Serge Laporte (IMT)

Mots-clé : Contrôle optimal, Méthode indirecte, Préconditionnement géométrique, Transformée de Mathieu, Interprétation géométrique du co-état, Véhicule hybride.

Résumé : Le principe du maximum de Pontryagin fournit des conditions nécessaires d'optimalité pour des problèmes de commande optimale en introduisant un co-vecteur associée à l'état, appelé co-état. En effet, la trajectoire d'état optimale doit être trouvée parmi les projections des trajectoires en état et co-état, appelées extrémales, données par le principe du maximum.

À partir d'une paire état et co-état initiale, il est possible, sous certaines hypothèses, de construire l'extrémale associée par intégration de la dynamique hamiltonienne. L'objectif de la méthode indirecte de tir est donc de rechercher le co-état au temps initial, et mène au calcul d'un zéro de la fonction de tir simple, ce qui est habituellement fait en utilisant un solveur de type Newton. La fonction de tir simple est connue pour être sensible à la condition initiale. De plus, une bonne initialisation doit être donnée pour s'assurer que le solveur converge.

On propose une méthode de préconditionnement de la fonction de tir [1], basée d'une part sur une interprétation géométrique du co-état en lien avec l'ensemble accessible du système augmenté, et d'autre part sur la transformée de Mathieu qui fournit un changement de variables dans l'espace des phases à partir d'un difféomorphisme sur l'état.

On applique cette nouvelle méthode sur le problème de répartition de couple d'un véhicule hybride électrique. Le préconditionnement est construit à partir d'une transformation linéaire d'une ellipse en un cercle. On montre numériquement que cette méthode permet de réduire le nombre d'itérations de notre solveur. De plus, dans nos expériences, il est préférable d'utiliser le préconditionnement plutôt que d'avoir une bonne initialisation de la fonction de tir obtenue à partir d'une approximation de la fonction valeur [2].

Références

- [1] O. Cots, R. Dutto, S. Laporte, S. Jan. Geometric preconditioner for indirect shooting and application to hybrid vehicle. 2024. (hal-04473962)
- [2] O. Cots, R. Dutto, S. Laporte, S. Jan. A bilevel optimal control method and application to the hybrid electric vehicle. 2023. (hal-04359870)

Stability and optimal control of the chemostat system including a mutation term

- **Claudia Alvarez** (Avignon Université, Laboratoire de Mathématiques d'Avignon (EA 2151), F-84018 Avignon, France. claudia.alvarez-latuz@univ-avignon.fr)
- **Térence Bayen** (Avignon Université, Laboratoire de Mathématiques d'Avignon (EA 2151), F-84018 Avignon, France. terence.bayen@univ-avignon.fr)
- **Jérôme Coville** (UR 546 Biostatistique et Processus Spatiaux, INRAE, Domaine St Paul Site Agroparc, F-84000 Avignon, France. jerome.coville@inrae.fr)

Mots-clé : Chemostat system, population dynamics, dynamical system, regular perturbation, global and local stability, optimal control

Résumé : The Chemostat system represents a specific class of dynamical systems describing the behavior of multiple species that compete over a same substrate within a reactor. This model finds application in diverse areas, notably in the investigation of wastewater treatment. Extensive literature exists on to the stabilization and optimal control of this system. Within these results, the competitive exclusion principle is a key finding, stating that the system attains a steady state where only one dominant species is present. However, empirical scenarios often reveal the presence of residual biomass from other species, a phenomenon that could be attributed to several phenomena and in particular to inter-species mutations. That is why, we consider here a modified version of the system incorporating a mutation term as a persistent perturbation. Our first objective is to establish that the system globally converges to a steady state that does not fulfill the competitive exclusion principle (*i.e.*, all species are present at steady-state), for any sufficiently small perturbation (in contrast with [1, 2] where this result was obtained provided that the dilution rate is small enough). Next, we wish to study an optimal control problem which amounts to maximize a biodiversity index (Shannon's index). The stability result provides a first insight into this problem by solving this optimization problem at steady state.

Références

- [1] T. Bayen, J. Coville, F. Mairet. Stabilization of the chemostat system with mutations and application to microbial production, to appear in *to appear in Optimal Control Appl. Methods*, <https://doi.org/10.1002/oca.3041>, 2023.
- [2] T. Bayen, H. Cazenave-Lacroutz, J. Coville. Stability of the chemostat system including a linear coupling between species, *Discrete and Continuous Dynamical Systems - Series B*, vol. 28,3, pp.2104–2129, 2023.

Estimating Signatures using ADMM for sparse third order tensor factorisation

- Stéphane Chrétien (Université Lyon 2 Laboratoire ERIC)
- Rémi Vaucher (Université Lyon 2 Laboratoire ERIC)

Mots-clé : Tensor estimation, Polynomial optimisation, Signature theory

Résumé : Signatures can be computed for multidimensional signals that are [irregularly sampled](#) in [unsynchronised ways across dimensions](#), a feature that make the approach very handy for many applications

The signature of order k ,

$$S_{[0,t]}^{(k)}(X) \in \mathbb{R}^{\overbrace{d \times d \times \dots \times d}^{k \text{ times}}} \quad (1)$$

of a path (in our case, a signal)

$$X(t) = (X(t)^1, \dots, X(t)^d) \quad (2)$$

$t \in [0, T]$ is defined for every word $i_1 i_2 \dots i_k$ from $\{1, \dots, d\}$ by

$$S(X)_{0,t}^{i_1, \dots, i_k} = \int_{0 < t_k < t} \dots \int_{0 < t_1 < t_2} dX_{t_1}^{i_1} \dots dX_{t_k}^{i_k}. \quad (3)$$

Signatures have very useful properties that make them ideal for feature extraction.

As recently demonstrated in [1], one can profitably use the decomposition of the signal in an appropriate basis, e.g. a wavelet basis into which the signal has a very sparse decomposition. For this purpose, define a (finite for computational simplicity) dictionary $\psi = (\psi_1, \dots, \psi_m) : [0, 1] \rightarrow \mathbb{R}^m$ and a linear map $A = (a_{ij}) \in \mathbb{R}^{n \times m}$ such that

$$X = A\psi : t \mapsto \left(\sum_{i=1}^m a_{1,i} \psi_i(t), \dots, \sum_{i=1}^m a_{n,i} \psi_i(t) \right).$$

It was proved in [1] that

$$S^{(3)}(X) = S^{(3)}(A\psi) = [[C_\psi; A, A, A]] \in \mathbb{R}^{n \times n \times n} \quad (4)$$

where $C_\psi = S^{(3)}(\psi) = (c_{ijk})_{i,j,k \in [1:n]}$ is the signature of the dictionary.

The goal of our work is to study the performance of ADMM for solving the estimation problem of $S^{(3)}(\mathbb{E}[X])$, i.e. the signature of the original signal, by solving the following least-square regression problem:

$$\min_{A \in \mathbb{R}^{n \times m}} \left\| S^{(3)}(X) - [[C; A, A, A]] \right\|_F^2$$

and its sparse regularisations involving the ℓ_1 norm. We demonstrate our results with promising numerical experiments.

Références

- [1] Pfeffer, M., Seigal, A., and Sturmfels, B. Learning paths from signature tensors. *SIAM Journal on Matrix Analysis and Applications*, 40(2), 394-416, 2023.

Projected gradient descent accumulates at Bouligand stationary points

- **Guillaume Olikier** (Université Côte d'Azur and Inria)
- Irène Waldspurger (CNRS, Université Paris Dauphine)

Mots-clé : low-rank optimization, convergence analysis, stationarity, first-order method, tangent cones, determinantal variety

Résumé : This work concerns the projected gradient descent (PGD) algorithm for the problem of minimizing a continuously differentiable function on a nonempty closed subset of a Euclidean vector space. Without further assumptions, this problem is intractable and devoted algorithms are only expected to find a stationary point. PGD is known to generate a sequence whose accumulation points are Mordukhovich stationary. In this work these accumulation points are proven to be Bouligand stationary, and even proximally stationary if the gradient is locally Lipschitz continuous. These are the strongest stationarity properties that can be expected for the considered problem.

Beyond the Fermat optimality rules

- Malek Abbasi (Department of Mathematics, University of Isfahan, Isfahan, Iran)
- **Sorin-Mihai Grad** (Unité de Mathématiques Appliquées, ENSTA Paris, 91120 Palaiseau, France)
- Michel Théra (Mathematics and Computer Science Department, University of Limoges)

Mots-clé : Fermat optimality conditions, Moreau-Rockafellar subdifferential, Sup-subdifferential, Outer/upper/inner-semicontinuity, Kuratowski-continuity, Hamel basis

Résumé : In this talk based on [1] we propose a general framework for analyzing the behavior at its extrema of an extended real-valued function that can lack convexity or differentiability, and for which the classical Fermat rules of optimality do not apply. To this end, we employ the notions of *sup-subdifferential*, recently introduced by two of the authors together with A. Kruger in [2], and *partial sup-subdifferentials*. The sup-subdifferential is a *nonempty* enlargement of the Moreau-Rockafellar subdifferential, satisfying most of its fundamental properties and enjoying certain calculus rules. The partial sup-subdifferentials are obtained by breaking down the sup-subdifferential into *one-dimensional components* through the elements of a Hamel basis and play the same role as the partial derivatives in the Fermat optimality rules.

Références

- [1] M. Abbasi, S.-M. Grad, M. Théra. Beyond the Fermat optimality rules. *in revision*.
- [2] M. Abbasi, A. K. Kruger, M. Théra. Enlargements of the Moreau-Rockafellar subdifferential. *Set-Valued and Variational Analysis*, 29(3):701–719, 2021.

Existence, Régularité et Calcul des potentiels duaux pour le transport optimal faible

- Guillaume Carlier (Université Paris-Dauphine PSL - CEREMADE)
- Maxime Sylvestre (Université Paris-Dauphine PSL - CEREMADE)
- Hugo Malamut (Université Paris-Dauphine PSL - CEREMADE)

Mots-clé : Transport optimal faible, Algorithme de Sinkhorn, Transport Martingale, Transport inverse

Résumé : Le transport optimal faible (WOT) [1] est une extension du transport optimal permettant d'englober des problèmes du type : transport martingale [3], régression de vecteur de quantile [2], interpolation de Brenier-Strassen [4], transport inverse [5]... Des résultats de dualité sont connus dans le cas non régularisé du problème WOT lorsque le coût est non dégénéré [1]. Récemment des avancées portant sur le transport martingale [6] démontrent l'existence de potentiels pour ce cas qui comporte un coût dégénéré du fait de la contrainte de martingale. Nous proposons ici un cadre général, dans le cas de marges à supports compacts dans \mathbb{R}^d , permettant d'assurer l'existence et la régularité (à minima L^∞) des potentiels duaux pour des problèmes de WOT avec ou sans régularisation entropique. Ce cadre inspiré d'une version forte du théorème de Strassen [7] inclut les différents types de coûts que l'on retrouve dans les applications ci-dessus: coût-faible Lipschitz sur l'espace des probabilités, contrainte/pénalisation de moments du noyau de transition et régularisation entropique. En sont déduits des résultats de stabilité par rapport aux marges, de Γ -convergence en les contraintes/pénalisation et de vitesse de convergence du problème ϵ -régularisé en $O(\epsilon \ln(1/\epsilon))$ vers le problème non régularisé. Pour le calcul pratique des solutions nous proposons une approche hybride mêlant algorithme de Sinkhorn et méthodes de 1er/2nd ordre. Nous démontrons une vitesse de convergence linéaire, observée en pratique.

Références

- [1] Gozlan, N., Roberto, C., Samson, P., Tetali, P., Kantorovich duality for general transport costs and applications, *Journal of Functional Analysis*, Volume 273, Issue 11, 2017.
- [2] Carlier, G., Chernozhukov, V., Galichon, A., Vector Quantile Regression: An Optimal Transport Approach, *Annals of Statistics*, Volume 44, No 3, June 2016
- [3] Beiglböck, M., Henry-Labordère, P., Penkner, F., Model-independent Bounds for Option Prices: A Mass Transport Approach, *Finance and Stochastics*, July 2013, Volume 17, Issue 3, pp 477–501
- [4] Gozlan, N. and Juillet, N. , On a mixture of Brenier and Strassen Theorems, *Proc. London Math. Soc.*, 120: 434-463, 2020
- [5] Dupuy, A. and Galichon, A., Personality traits and the marriage market, *Journal of Political Economy*, 122(6):1271–1319, 2014
- [6] Backhoff-Veraguas, J., Beiglböck, M., Schachermayer, W., Tschiderer, B., The structure of martingale Benamou–Brenier in \mathbb{R}^d , arXiv, 2306.11019, 2023
- [7] Strassen, V., The Existence of Probability Measures with Given Marginals, *Ann. Math. Statist.* 36(2): 423-439 (April, 1965)

Non-linear and non-convex sparse coding for DNA analysis in single molecule

- **Clara Lage** (Univ Lyon, ENS de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France)
- Nelly Pustelnik (Univ Lyon, ENS de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France)
- Benjamin Audit (Univ Lyon, ENS de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France)
- Jean-Michel Arbona (Laboratoire de Biologie et Modélisation de la Cellule, ENS de Lyon, Lyon, France)
- Rémi Gribonval (Univ Lyon, Inria, EnsL, UCBL, CNRS, LIP, UMR 5668, F-69342)

Mots-clé : Non-convex optimization, Non-linear inverse problems, Sparse coding, DNA replication analysis

Résumé : The DNA synthesis, also known as DNA replication, is the process by which a cell doubles its DNA content in order to have two identical copies of its chromosomes. This crucial process ensures that genetic information is accurately passed on to the next generation of cells. One of the current objectives in DNA replication research is to estimate the starting points and local speeds of the molecular motors responsible for the DNA synthesis along chromosomes. This can be achieved by the FORK-seq technique [2], that uses DNA sequencing to generate a signal from which the molecular movement can be accurately detected with a resolution close to a base pair.

In previous studies, numerical methods applied to FORQ-seq data successfully enabled the positioning, orientating, and extraction of the velocity information of molecular motors involved in DNA synthesis [3, 2]. Nevertheless, a substantial part of this data remains unexplored due to the fact that the model is not additive: signal patterns associated to two molecular motors overlap.

In this work, we provide a new formulation of the movement identification problem to overcome this methodological limitation. The new optimization problem is formulated as a *nonlinear sparse coding problem* in the set of piecewise linear vectors representing the local constant velocity movement of molecular motors. We propose an optimization scheme capable of handling the entire dataset provided by the FORK-seq high-throughput experimental technique, and compare it with a well-known primal-dual proximal optimization method [1].

Références

- [1] T. Valkonen. Acceleration and Global Convergence of a First-Order Primal-Dual Method for Nonconvex Problems. *SIAM Journal on Optimization*, 93:1–42, 2021.
- [2] B. Theulot and al. Genome-wide mapping of individual replication fork velocities using nanopore sequencing. *Nature Communications*, 13:125–139, 2022.
- [3] C. Lage, N. Pustelnik, J.-M. Arbona and B. Audit. Codage espace-échelle parcimonieux en présence de bruit non-gaussien. Application à l’analyse de la réplication de l’ADN en molécule unique. *Gretsi*, 2023.

A second-order accurate space-time discretization for L^2 -Wasserstein gradient flows with exact energy dissipation

- Clément Cancès (Inria, Université de Lille)
- Léonard Monsaingeon (University of Lisbon)
- **Andrea Natale** (Inria, Université de Lille)

Mots-clé : Wasserstein gradient flows, Fokker-Planck equation, TPFA finite volumes

Résumé : In this talk we discuss the construction and numerical analysis of a space-time discretization for L^2 -Wasserstein gradient flows for energies of the form:

$$\mathcal{E}(\rho) = \int_{\Omega} E(\rho) + V\rho,$$

where $E : [0, \infty) \rightarrow \mathbb{R}$ is a convex function and $V : \Omega \rightarrow \mathbb{R}$ is a given potential. Under appropriate assumptions, one can characterize such gradient flows as the absolutely continuous curves $\rho : [0, T] \rightarrow \mathcal{P}(\Omega)$ verifying the energy dissipation equality,

$$\mathcal{E}(\rho(t)) + \frac{1}{2} \int_0^t |\rho'|^2 dr + \frac{1}{2} \int_0^t |\partial\mathcal{E}|(\rho(r))^2 dr = \mathcal{E}(\rho(0)), \quad \forall t \in [0, T], \quad (1)$$

where $|\rho'|$ and $|\partial\mathcal{E}|$ denote, respectively, the metric derivative of the curve and the slope of \mathcal{E} , with respect to L^2 -Wasserstein metric. We construct a numerical scheme which verifies exactly a discrete version of this relation. The main idea is to use an appropriately defined nonlinear mean between the discrete solution at two successive times, coupled with the square-root approximation finite volume scheme for the space discretization [1]. In the case of $E(r) = r \log r - r$, which gives rise to the Fokker-Planck equation, we prove that the Γ -lim inf of our discrete functional replacing (1) is bounded from below by its continuous counterpart, which immediately implies convergence of the scheme. Finally we propose an efficient numerical scheme to compute the discrete solutions of our scheme and show numerically second-order convergence with respect to both space and time discretization parameters.

Références

- [1] C. Cancès and J. Venel. On the square-root approximation finite volume scheme for non-linear drift-diffusion equations. *Comptes Rendus. Mathématique*, 361.G2: 535-558, 2023.

Transport optimal partiel et estimation de risque

- **Adrien Cances** (Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay)

Mots-clé : transport optimal partiel, estimation de risque, méthodes particulières

Résumé :

Ce travail est motivé par un problème d'estimation de risque [1] qui consiste à trouver la loi jointe (les marginales étant fixées) maximisant le α -risque, qui désigne la moyenne de la proportion α des plus grandes valeurs de la mesure image de la loi jointe par une certaine fonction dite *fonction de risque*. Ce problème d'estimation de risque admet naturellement deux formulations. La première consiste en un problème de transport optimal multi-marges standard, à ceci près que la fonction objectif n'est plus linéaire. La seconde formulation donne quant à elle un problème de transport optimal multi-marges où l'objectif est bien linéaire, mais où on ne transporte qu'une fraction α de la masse.

Des résultats sur le transport optimal multimarge suggèrent que la loi jointe est supportée par un ensemble de petite dimension, au moins lorsque la fonction de risque est supermodulaire. L'idée principale de ce travail consiste en une formulation *particulière* du problème d'estimation de risque : on cherche une solution dans l'ensemble des lois à support discret fini, paramétrisé par la position des masses de Dirac. Les contraintes de marge ne pouvant pas être vérifiées exactement par ces lois, on les transforme en contraintes douces via une pénalisation utilisant des distances de Wasserstein 1D, standards ou partielles, selon la formulation utilisée. Nous introduisons un algorithme efficace pour résoudre le transport partiel 1D, un problème plus difficile qu'il n'y paraît, et qui pourrait avoir un intérêt dans d'autres contextes [2].

Références

- [1] Mérigot, Quentin and Nenna, Luca and Ennaji, Hamza and Pass, Brendan Robust risk management via multi-marginal optimal transport *arXiv preprint arXiv:2208.05888*, 2022.
- [2] Nicolas Bonneel and David Coeurjolly, *Spot: sliced partial optimal transport*, *ACM Transactions on Graphics (TOG)*, **38**(4), 1–13, 2019, ACM New York, NY, USA.

Optimisation de forme et inéquations variationnelles : le problème de Tresca dans le modèle 2D de l'élasticité linéaire

- Loïc Bourdin (Université de Limoges, Institut de recherche XLIM, UMR CNRS 7252, 87000 Limoges, France)
- Fabien Caubet (Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, UMR 5142, 64000 Pau, France)
- **Aymeric Jacob de Cordemoy** (Sorbonne Université, CNRS, Université Paris Cité, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France)

Mots-clé : optimisation de forme, loi de Tresca, conditions unilatérales de Signorini, mécanique du contact, opérateur proximal, épi-différentiabilité d'ordre 2

Résumé : Dans ce travail nous étudions, sans aucune procédure de régularisation ou de pénalisation, un problème d'optimisation de forme impliquant un phénomène de frottement dans le modèle 2D de l'élasticité linéaire. Ce frottement est modélisé par la *loi de Tresca* (e.g., [2]) qui apparaît comme une condition au bord impliquant des inégalités et des termes non lisses. Plus précisément, nous considérons un problème aux limites où intervient cette loi, le problème de Tresca, et nous en effectuons *l'analyse de sensibilité* par rapport à la forme à l'aide de la théorie des inéquations variationnelles, d'outils d'analyse convexe et non lisse, tels que *l'opérateur proximal* [4] ou encore la notion *d'épi-différentiabilité du second ordre* [5]. Nous prouvons alors que la solution du problème de Tresca admet une dérivée directionnelle par rapport à la forme, qui coïncide avec la solution d'un problème aux limites impliquant des *conditions tangentielles de Signorini* (e.g., [3, Section 1]). Ensuite, le *gradient de forme* (e.g., [1]) de la fonctionnelle énergie de Tresca est explicitement caractérisé, nous permettant d'effectuer des simulations numériques.

Références

- [1] A. Henrot et M. Pierre. Shape Variation and Optimization : a Geometrical Analysis. *Tracts in Mathematics Vol. 28, European Mathematical Society*, 2018
- [2] F. Kuss. Méthodes duales pour les problèmes de contact avec frottement. *Université de Provence - Aix-Marseille I*, July 2008.
- [3] J.-L. Lions. Sur les problèmes unilatéraux. *Séminaire Bourbaki, Springer-Verlag*, : vol. 1968/69, exposés 347-363, 1971.
- [4] J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.
- [5] R. T. Rockafellar. Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, Vol. 2, p. 167–184, 1985.

Is the Faber-Krahn inequality true for the Stokes operator?

- Antoine Henrot (IECL, Université de Lorraine)
- **Idriss Mazari-Fouquer** (Ceremade, Paris Dauphine Université PSL)
- Yannick Privat (IECL, Université de Lorraine)

Mots-clé : Shape optimisation, Spectral optimisation, Stokes operator

Résumé : In this work in collaboration with A. Henrot and Y. Privat, we investigate the minimisation of the first eigenvalue of the Dirichlet-Stokes operator in \mathbb{R}^d . To be more specific, we consider the question of existence of optimal shapes, a query made more difficult by the incompressibility constraint, as well as the minimality of the ball. This latter point turns out to be dependent on the dimension, and starkly different from the case of the (scalar) Dirichlet-Laplacian.

Références

- [1] A. Henrot, I. Mazari-Fouquer, Y. Privat, Is the Faber-Krahn inequality true for the Stokes operator? *Submitted, 2024.*

Don't choose between workload balance and makespan minimization

- Sébastien Deschamps (Saint-Gobain Recherche)
- Frédéric Meunier (École des Ponts)

Mots-clé : bipartite graph, load balancing, fairness

Résumé : In a simple model of assigning workers to tasks, every solution that minimizes the load difference between the most loaded worker and the least loaded one actually minimizes the maximal load and maximizes the minimal load. This can be seen as a consequence of standard results on lexicographically optimal bases of polymatroids (which go back to the works of Meggido [2] and Fujishige [1]). We show that similar phenomena still occur in close models, simple to state, and that do not enjoy any polymatroid structure.

Références

- [1] Satoru Fujishige. Lexicographically optimal base of a polymatroid with respect to a weight vector. *Mathematics of Operations Research*, 5:186–196, 1980.
- [2] Nimrod Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming*, 7:97–107, 1974.

La distance entre polytopes en nombres entiers

- Antoine Deza (McMaster University, Hamilton, Ontario, Canada)
- Shmuel Onn (Technion – Institut de technologie d’Israël, Haifa, Israël)
- Sebastian Pokutta (Zuse-Institute Berlin, Allemagne)
- **Lionel Pournin** (Université Paris 13, Villetaneuse, France)

Mots-clé : projections alternées, polytopes en nombres entiers, optimisation

Résumé : Considérons deux polytopes en nombres entiers disjoints contenus dans l’hypercube $[0, k]^d$. Puisqu’il y a un nombre fini de telles paires, la plus petite distance possible $\varepsilon(d, k)$ entre ces deux polytopes est un nombre positif. Ce nombre apparait dans la complexité de l’algorithme de projections alternées de Von Neumann [3]. Il est également lié aux notions de *distance faciale* [5, 7, 8], de *distance sommet-facette* [2] et de *largeur pyramidale* [6, 9]. Des bornes ont été données par Noga Alon et Văn Vū sur la plus petite distance sommet-facette possible d’un simplexe en nombres binaires [1].

Nous obtenons [4] des bornes sur $\varepsilon(d, k)$ qui montrent en particulier que cette quantité diminue de façon exponentiellement rapide avec d et k . Il en découle des bornes sur la distance faciale des polytopes en nombres entiers similaires à celles données par Noga Alon et Văn Vū dans le cas de la distance sommet-facette des simplexe en nombres binaires. Ces résultats permettent aussi de borner la distance minimale entre deux polytopes en nombres rationnels disjoints en fonction de la taille combinée de leur représentation sous forme de système d’inégalités linéaires.

Références

- [1] Noga Alon et Văn H. Vū, *Anti-Hadamard matrices, coin weighing, threshold gates, and indecomposable hypergraphs*, Journal of Combinatorial Theory, Series A **79** (1997), no. 1.
- [2] Amir Beck et Shimrit Shtern, *Linearly convergent away-step conditional gradient for non-strongly convex functions*, Mathematical Programming **164** (2017), no. 1–2, 1–27.
- [3] Gábor Braun, Sebastian Pokutta et Robert Weismantel, *Alternating linear minimization: revisiting von Neumann’s alternating projections*, arXiv:2212.02933 (2022).
- [4] Antoine Deza, Shmuel Onn, Sebastian Pokutta et Lionel Pournin, *Kissing polytopes*, arXiv:2305.18597 (2023).
- [5] David H. Gutman et Javier F. Peña, *The condition number of a function relative to a set*, Mathematical Programming **188** (2021), no. 1, 255–294.
- [6] Simon Lacoste-Julien et Martin Jaggi, *On the global linear convergence of Frank–Wolfe optimization variants*, Proc. of NIPS 2015, pp. 496–504.
- [7] Javier F. Peña, *Generalized conditional subgradient and generalized mirror descent: duality, convergence, and symmetry*, arXiv:1903.00459 (2019).
- [8] Javier F. Peña et Daniel Rodríguez, *Polytope conditioning and linear convergence of the Frank–Wolfe algorithm*, Mathematics of Operations Research **44** (2018), no. 1, 1–18.
- [9] Luis Rademacher et Chang Shu, *The smoothed complexity of Frank–Wolfe methods via conditioning of random matrices and polytopes*, Mathematical Statistics and Learning **5** (2022), no. 3/4, 273–310.

Interior Point Methods with a Gradient Oracle

- **Adrian Vladu** (CNRS, IRIF, Université Paris Cité)

Mots-clé : interior point methods, linear systems, preconditioning

Résumé : Gradient methods are popular optimization primitives, since they rely entirely on first-order information, which is usually easy to obtain. Nevertheless, many interesting problems, such as those appearing in conic optimization involve hard constraints, which require the use of more sophisticated techniques (e.g. cutting planes or interior point methods). These typically involve exploiting higher order information about the objective or the geometry of the domain, such as the Hessian matrix.

In this talk, I will show how to overcome this difficulty, by efficiently executing an interior point method which does not have direct access to second order derivatives, and is solely based on gradient information.

The key ingredient of this method is a beautiful idea of [Dunagan-Harvey '07], where it was shown how to solve linear systems of equations while simultaneously constructing a preconditioner. This approach can be leveraged in the context of interior point methods, where one needs to solve a sequence of linear systems with a slowly changing matrix.

As a consequence, we can compute high precision solutions to linear optimization problems over (reasonably well-conditioned) n -dimensional convex bodies using $\tilde{O}(n)$ gradient queries to an appropriate barrier function, plus an extra cubic runtime overhead.

Un modèle de Cahn-Hilliard d'ordre deux pour l'approximation de flots de diffusion de surface de structures fines

- Elie Bretin (ICJ-INSA)
- Luc Calatroni (I3S-CNRS)
- Roland Denis (ICJ-CNRS)
- Simon Masnou (ICJ-Lyon 1)
- Garry Terii (ICJ-Lyon 1)
- Arnaud Sengers (ICJ-Lyon 1)

Mots-clé : Cahn Hilliard, Diffusion de surface, discrétisation numérique

Résumé : Nous nous intéressons, dans cet exposé, à l'approximation de flots de diffusion de surface à l'aide d'un modèle de type Cahn-Hilliard. Plus précisément et afin de compenser des pertes de masse observée dans l'approximation d'évolution de structures fines, nous proposons un nouveau modèle variationnel d'ordre deux obtenu comme le flot de gradient de l'énergie classique de Cahn-Hilliard, mais suivant une topologie composée de deux mobilités dégénérées [2, 3]. Nous expliquerons également comment adapter des approches SAV [1] pour traiter numériquement ces mobilités et obtenir des schémas d'approximation simples et efficaces. Des simulations numériques 3D du mouillage d'un tube fin sur différents supports solides illustreront l'efficacité de ce nouveau modèle.

Références

- [1] E. Bretin, L. Calatroni and S. Masnou, A mobility-SAV approach for a Cahn-Hilliard equation with degenerate mobilities. *Discrete and Continuous Dynamical Systems Series S*, to appear, 2023
- [2] E. Bretin, R. Denis, S. Masnou, A. Sengers, and G. Terii, A Cahn-Hilliard multiphase system with mobilities for the simulation of wetting *ESAIM: M2AN (Mathematical Modelling and Numerical Analysis)*, 57(3):1473-1509, (2023)
- [3] E. Bretin, S. Masnou, A. Sengers and G. Terii, Approximation of surface diffusion flow: a second order variational Cahn-Hilliard model with degenerate mobilities *M3AS*, vol 32, 04, 793-829 (2022)

Trading-off price for data quality to achieve fair online allocation

- Mathieu Molina (Inria, FairPlay Team, Palaiseau, France)
- Nicolas Gast (Univ. Grenoble Alpes, Inria, CNRS)
- Patrick Loiseau (Inria, FairPlay Team, Palaiseau, France)
- **Vianney Perchet** (CREST, ENSAE & Criteo AI Lab, France)

Mots-clé : Fairness, Privacy, Bandits

Résumé : We consider the problem of online allocation subject to a long-term fairness penalty. Contrary to existing works, however, we do not assume that the decision-maker observes the protected attributes—which is often unrealistic in practice. Instead, they can purchase data that help estimate them from sources of different quality; and hence reduce the fairness penalty at some cost. We model this problem as a multi-armed bandit problem where each arm corresponds to the choice of a data source, coupled with the online allocation problem. We propose an algorithm that jointly solves both problems and shows that it has a regret bounded by $\mathcal{O}(\sqrt{T})$. A key difficulty is that the rewards received by selecting a source are correlated by the fairness penalty, which leads to a need for randomization (despite a stochastic setting). Our algorithm takes into account contextual information available before the source selection and can adapt to many different fairness notions. We also show that in some instances, the estimates used can be learned on the fly.

Références

- [1] Molina, M., Gast, N., Loiseau, P. and Perchet, V.

Trading-off price for data quality to achieve fair online allocation, *NeurIPS*, 2023

Open Problems about the Simplex Method

- **Sophie Huiberts** (CNRS, LIMOS)

Mots-clé :

Résumé : The simplex method is a very efficient algorithm. In this talk we see a few of the state-of-the-art theories for explaining this observation. We will discuss what it takes for a mathematical model to explain an algorithm's qualities, and whether existing theories meet this bar.

From Conformal Predictions to Confidence Regions

- Eugene Ndiaye (Apple)

Mots-clé :

Résumé : If you predict a label y of a new object with \hat{y} , how confident are you that “ $y = \hat{y}$ ”? The conformal prediction method provides an elegant framework for answering such a question by establishing a confidence set for an unobserved response of a feature vector based on previous similar observations of responses and features. This is performed without assumptions about the distribution of the data. While providing strong coverage guarantees, computing conformal prediction sets requires adjusting a predictive model to an augmented dataset considering all possible values that the unobserved response can take, and proceeding to select the most likely ones. For a regression problem where y is a continuous variable, it typically requires an infinite number of model fits; which is usually infeasible. By assuming a little more regularity in the underlying prediction models, I will describe some of the techniques that make the calculations feasible. Along similar lines, it can be assumed that we are working with a parametric model that explains the relation between input and output variables. Consequently, a natural question arises as to whether a confidence interval on the ground truth parameter of the model can be constructed, also without assumptions on the distribution of the data. In this presentation, I will provide some preliminary results and discuss remaining open questions.

Références

- [1] Molina, M., Gast, N., Loiseau, P. and Perchet, V.
Trading-off price for data quality to achieve fair online allocation, *NeurIPS*, 2023

L'optimisation au service de la transition énergétique : quelques avancées récentes

- Sandrine Charoussat-Brignol (EDF R&D)

Mots-clé :

Résumé : Les objectifs européens de réduction des émissions nécessitent des changements profonds du système électrique : forte croissance de la demande en électricité, pénétration de plus en plus importante des énergies décarbonées, en particulier renouvelables. . . . Le système électrique devient de plus en plus complexe à modéliser. Alors que par le passé il s'agissait d'équilibrer une consommation peu flexible avec une production basée sur des centrales contrôlables (centrales thermiques charbon, gaz, nucléaire, et centrales hydrauliques principalement), il faut aujourd'hui intégrer dans la modélisation de nombreux éléments supplémentaires : la production à base d'éolien ou de solaire représente une part importante, et son caractère volatile ainsi que sa dépendance aux conditions météorologiques induisent une très forte augmentation du besoin de flexibilité afin de pouvoir équilibrer le système ; de nouveaux moyens de stockage (batteries, voitures électriques avec 'smart-charging'. . . .) ainsi que des consommateurs qui deviennent acteurs du système. Nous présenterons dans cet exposé des éléments de modélisation permettant de prendre en compte ces évolutions, ainsi que des avancées méthodologiques permettant de résoudre les grands problèmes qui y sont attachés.