Two-sided matching markets with correlated random preferences have few stable pairs

Claire Mathieu
CNRS

Stable matching in a community consisting of N men and N women is a classical combinatorial problem that has been the subject of intense theoretical and empirical study since its introduction in 1962 in a seminal paper by Gale and Shapley.
In this paper, we study the number of stable pairs, that is, the man/woman pairs that appear in some stable matching. We prove that if the preference lists on one side are generated at random using the popularity model of Immorlica and Mahdian, the expected number of stable edges is bounded by $N \ln(N) + N$, matching the asymptotic value for uniform preference lists. If in addition that popularity model is a geometric distribution, then the number of stable edges is $O(N)$ and the incentive to manipulate is limited. If in addition the preference lists on the other side are uniform, then the number of stable edges is asymptotically N up to lower order terms: most participants have a unique stable partner, hence non-manipulability.
This is joint work with Hugo Gimbert and Simon Mauras.

Optimisation convexe et non-convexe pour l'apprentissage statistique

Francis Bach
Inria et ENS

L'apprentissage statistique se formule naturellement comme un problème d'optimisation, qui cherche à minimiser le taux d'erreur sur les données observées, typiquement de grande taille. Si pour les modèles linéaires, la fonction objectif est convexe, elle ne l'est pas pour des modèles non-linéaires comme les réseaux de neurones. Dans cet exposé, je présenterai certains développements récents liés à ces deux situations.

Some optimization issues in the new electrical system

Nadia Oudjane
EDF-Lab

With the emergence of new renewable energies (as wind or solar generation), the electrical system is evolving from a situation based on controllable power plants (thermal, gas, hydro, etc.) connected to the transmission network to a new setting, where substantial uncontrollable generation facilities are connected to the distribution network, closer to the consumers. To contribute to this transformation, new local actors are rapidly multiplying potentially involving a great variety of devices: uncontrollable renewable generations; storage devices (batteries, Electric Vehicles smart charging systems, ...); conventional plants (such as gas turbines or hydro plants); possibilities to adjust some consumers load (e.g. demand response, or direct control). In this new context, energy utilities are facing new practical issues requiring to develop suitable mathematical optimization tools.

How to discretize the total variation?

Antonin Chambolle
CMAP, CNRS, École polytechnique, IP Paris

This talk will address standard and less standard approaches for discretizing the total variation, widely used as a regularizer for solving inverse problems in imaging. It is used for recovering functions with discontinuities (edges) and therefore standard numerical analysis, usually based on the smoothness of the solutions, yields sub-optimal error bounds for such problems. The talk will address some workarounds, discussing primal and dual discretizations, isotropy issues, error bounds. (Based on joint works with C. Caillaud and T. Pock.)

Sur la convergence de la méthode du gradient hybride primal-dual stochastique

Olivier Fercoq
LTCI, Télécom Paris, Institut Polytechnique de Paris

Joint work with *Ahmet Alacaoglu* and *Volkan Cevher*

Keywords : Convergence linéaire, descente par coordonnée, méthode primale-duale, sous-régularité métrique

Nous analysons la méthode du gradient hybride primal-dual stochastique (SPDHG) proposée récemment dans [1] et démontrons de nouveaux résultats théoriques. L'algorithme a pour but de résoudre les problèmes d'optimisation convexes du type

$$\min_{x \in \mathcal{X}} \sum_{i=1}^{n} f_i(A_i x) + g(x) \ .$$

À chaque itération, une unique coordonnée du vecteur de variables duales est mis à jour alors que toutes les coordonnées du vecteur primal sont modifiées.
Nous faisons l'hypothèse que le gradient généralisé du lagrangien est métriquement sous-régulier. Cette hypothèse de travail est satisfaite pour les problèmes d'optimisation linéaire, le Lasso, les Séparateurs à Vaste Marge et tous les problèmes lisses et fortement convexes.
Nous pouvons ainsi démontrer la convergence linéaire de la méthode. Fait notable, l'algorithme ne dépend pas de constante de sous-régularité métrique, ce qui explique qu'on observe la convergence linéaire même si on n'est pas capable d'estimer cette constante.
Dans le cas général, nous prouvons la convergence presque sûre des itérés vers une solution du problème et une vitesse de convergence sous-linéaire.
Enfin, nos expériences numériques montrent que SPDHG avec les pas standards a une bonne performance et est robuste comparé à la variante SPDHG-μ spécialisée pour les fonctions fortement convexes, au gradient stochastique avec réduction de variance et à la descente par coordonnée duale.

References

[1] Antonin Chambolle, Matthias J Ehrhardt, Peter Richtárik, and Carola-Bibiane Schonlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. *SIAM Journal on Optimization*, 28(4):2783–2808, 2018.

Pontryagin maximum principle for optimal sampled-data control problems

Gaurav Dhar
Université de Limoges

Joint work with *Loïc Bourdin*

Keywords : Optimal control, Pontryagin maximum principle, sampled-data control, optimal sampling times, Hamiltonian continuity, running state constraints, bouncing trajectory phenomenon.

Established in the 1950s by Pontryagin *et al.*, the Pontryagin maximum principle (in short, PMP) is a fundamental result in optimal control theory. It gives first-order necessary conditions for the optimal control among all *permanent* controls, that is to say when the value of the control can be changed at any moment in time. However, in practice the optimal permanent control of a given system is in general unrealizable. For this reason sampled-data controls, for which the value of the control can only be modified a finite number of times, are often implemented in Engineering and Automation. Recently Bourdin and Trélat have obtained in [1] a version of the PMP for optimal sampled-data control problems with fixed sampling times.

The work [1] does not take into account the possibility of *free sampling times*, where one is allowed to optimize the sampling times when the control can be modified in addition to the control values. We present a new version of the PMP for optimal sampled-data control problems with free sampling times where the additional necessary condition for the optimal sampling times coincides with the continuity of the Hamiltonian function (see [2]). Recall that this property is a well-known fact for optimal permanent controls. Therefore, our result asserts that the continuity property is recovered in the case of optimal sampled-data controls with optimal sampling times and it can be used, for example in shooting methods, to determine the optimal sampling times.

Finally we give a new version of the PMP for optimal sampled-data control problems with running state constraints (see [3]). The statement of the PMP is more involved since the adjoint vector is now described as a function of bounded variations. However, in the case of sampled-data controls, we find that, under certain general hypotheses, the optimal trajectory only contacts the running inequality state constraints at most at the sampling times also known as *the bouncing trajectory phenomenon*. Thus the adjoint vector only experiences jumps at most at the sampling times (and thus in a finite number and at precise instants) and its singular part vanishes.

References

[1] L. Bourdin, E. Trélat, *Optimal sampled-data control, and generalizations on time scales*, Mathematical Control and Related Fields, 2016.

[2] L. Bourdin, G. Dhar, *Continuity/constancy of the Hamiltonian function in a Pontryagin maximum principle for optimal sampled-data control problems with free sampling times*, Mathematics of Control, Signals, and Systems, 2019.

[3] L. Bourdin, G. Dhar, *Optimal sampled-data control problems with running state constraints – Pontryagin maximum principle and bouncing trajectories phenomenon*, submitted for publication.

Convergence Analysis of Single-call Extra-Gradient Methods

Yu-Guan Hsieh
LJK, Univ. Grenoble Alpes

Keywords : Variational Inequality, Extra-Gradient, Equilibrium

Variational inequalities (VIs) offer a flexible framework for modeling equilibrium problems, with applications ranging from economics, game theory to machine learning, among others. In this setting, the optimal $O(1/t)$ rate is achieved by the Extra-Gradient (EG) algorithm of Korpelevich [3] when the involved operator is monotone. However, EG requires two oracle calls per iteration, incurring an extra cost that can be significant in some applications. Aiming to alleviate this cost, several algorithms have been proposed as surrogates of EG with only a *single* oracle call per iteration, with notably the example of Popov's modified Arrow-Hurwicz method [4].

In this talk, we *(i)* develop a synthetic view of such *Single-call Extra-Gradient* (1-EG) algorithms, and *(ii)* show that they enjoy similar convergence guarantees as the standard EG algorithm for both deterministic and stochastic problems. Most importantly, we prove in [2] that the last iterate of 1-EG methods enjoys $O(1/t)$ local convergence to solutions of *stochastic*, *non-monotone* problems satisfying a second-order sufficient condition. The algorithms that are covered include the optimistic gradient method [1], the reflected gradient method and naturally Popov's modified Arrow-Hurwicz method [4].

	Monotone	Strongly Monotone	
	Ergodic	Ergodic	Last Iterate
Deterministic	$\boxed{1/t}$	$1/t$	$e^{-\rho t}$
Stochastic	$1/\sqrt{t}$	$\boxed{1/t}$	$\boxed{1/t}$

Table 1: The best known global convergence rates for EG and 1-EG methods for smooth operators. A box indicates our contribution conerning the analysis of single-call methods.

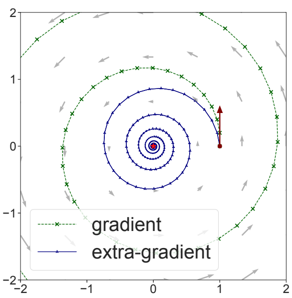


Figure 1: EG is the de facto standard algorithm for monotone VIs as gradient diverges in a simple bilinear game $\min_\theta \max_\phi \theta\phi$.

References

[1] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *ICLR*, 2018.

[2] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. In *NeurIPS*, pages 6936–6946, 2019.

[3] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12:747–756, 1976.

[4] Leonid Denisovich Popov. A modification of the Arrow–Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

a Randomized Proximal Gradient Method with Structure-Adapted Sampling

Franck Iutzeler
Univ. Grenoble Alpes

Keywords : Coordinate Descent, Nonsmooth Optimization, Identification

Many applications in machine learning involve nonsmooth optimization problems. This nonsmoothness brings a low-dimensional structure to the optimal solutions. In this talk, based on [3], we propose a randomized proximal gradient method harnessing this underlying structure.

Adaptive subspace descent Our problem writes

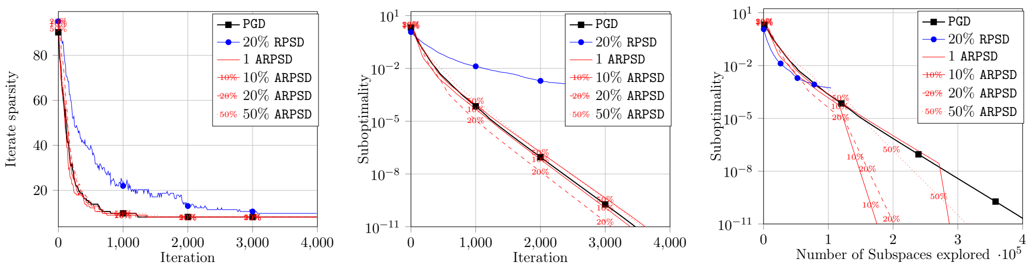
$$\min_{x\in\mathbb{R}^d} f(x) + g(x)$$

where f is a smooth strongly convex function and g is convex possibly non-differentiable. Let us examine the algorithm on the right:

for $k = 1, \ldots$ **do**
  $y^k = \mathsf{Q}\left(x^k - \gamma\nabla f\left(x^k\right)\right)$
  $z^k = P_{\mathfrak{S}^k}\left(y^k\right) + (I - P_{\mathfrak{S}^k})\left(z^{k-1}\right)$
  $x^{k+1} = \mathbf{prox}_{\gamma g}\left(\mathsf{Q}^{-1}\left(z^k\right)\right)$
end for

• if $P_{\mathfrak{S}^k} = I$ and $\mathsf{Q} = I$: this algorithm boils down to the proximal gradient method `PGD`.
• if $P_{\mathfrak{S}^k}$ is an i.i.d. sequence of projections and $\mathsf{Q} = \mathbb{E}[P_{\mathfrak{S}^1}]^{-\frac{1}{2}}$: this algorithm leads to a randomized *subspace descent* method `RPSD` for which we prove linear convergence.
• if $P_{\mathfrak{S}^k}$ is *adapted* to the *structure* of x^k (for instance, its sparsity). Then, we can show that if we do not allow too harsh changes, we obtain a linearly converging randomized algorithm `ARPSD` that automatically adapts to the structure of the iterates.

Applications Coordinate descent methods are very popular for large machine learning problems [1] but were limited to separable nonsmooth terms until recently [2]. In addition, adaptive randomized methods were almost non-existing and only heuristic, our work is the first to consider adaptive randomized descent for general composite problems backed by theory.
Let us consider the problem of logistic regression with 1D Total Variation regularization. The obtained iterates and solutions are thus constant by blocks and thus sparse by *jumps*. We see below that adaptively projecting the updates onto sets with the same block structure as the iterates brings a computational advantage in terms of directions explored.



References

[1] Y. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM Journal on Optimization (2012).

[2] F. Hanzely, K. Mishchenko, P. Richtarik, *Sega: Variance reduction via gradient sketching*, NeurIPS (2018).

[3] D. Grishchenko, F. Iutzeler, J. Malick, *a Proximal Gradient method with Adaptive Subspace Sampling*, to appear (2020).

A lagrangian discretization for mean field games with congestion:

Clément SARRAZIN
Institute of Mathematics of Orsay

Joint work with *F. SANTAMBROGIO and Q. MERIGOT*

Keywords : Optimization, Mean Field Games, Optimal Transport

Variationnal mean field games are differential games where an infinite number of players minimize a global energy accumulated during their movement inside a domain. Although these games usually arise in [2] from finite differential games as the number of players go to infinity, one can also consider terms in the energy depending on the density of the population (and have therefore no meaning for a finite population). These terms usually force players to move away from "congested zones", where the density is too high. Optimal strategies for these mean field games are often numerically found in a "eulerian way", by approximating the optimal densities at a finite number of time steps, and propagating from time step to time step, which necessitates an ad-hoc time discretization of the transport equation verified by the mean field game, and can be challenging.

The approach I would like to talk about is a more "Lagrangian" one, where one approximate the behaviour of the population with the strategies of a finite number of players. The main hinderance here is the term penalizing congestion, which is only finite when the population admits densities at all time. A way to circumvent this problem is to replace it with one essentially penalizing the distance to the set of measures not congested. This is inspired by a similar work on the incompressible Euler equation in [3]. Time discretization is simply done by considering only piecewise affine movements for the players.

I would give a definition for this term in a general case, using a Moreau-Yosida approximation for the Wasserstein distance. It can be numerically computed, by solving a smooth concave problem, with computations very reminiscent of the ones in [1] for semi-discrete optimal transport .
Such a term defines an energy, finite at discrete measures, and the optimal, discrete, strategies for this new game converge in some sense to optimal strategies for the initial, continuous, mean field game. I would also show a more "observable" convergence for the Moreau-Yosida approximations and some numerical simulations to go with it.

References

[1] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *arXiv preprint arXiv:1603.05579*, 2016.

[2] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, Mar 2007.

[3] Quentin Mérigot and Jean-Marie Mirebeau. Minimal geodesics along volume-preserving maps, through semidiscrete optimal transport. *SIAM Journal on Numerical Analysis*, 54(6):3465–3492, 2016.

Parameter Estimation for Dynamic Resource Allocation in Microorganisms: A Bi-level Optimization Problem

Térence Bayen
University of Avignon

Joint work with *Francis Mairet*

Keywords : Optimal control, Bi-level optimization, Microbial growth, Pontryagin's Principle

Given their key role in almost all ecosystems and in several industries, understanding and predicting microorganism growth is of utmost importance [1]. In compliance with evolutionary principles, coarse-grained or genome-scale models of microbial growth can be used to determine optimal resource allocation scheme under dynamic environmental conditions. Resource allocation have given important qualitative results, but it still remains a gap towards quantitative predictions. The first step in this direction is parameter calibration with experimental data.
In this work, we present a coarse-grained model describing how microalgae acclimate to a change in light intensity. Fitting this model results in a bi-level optimization problem,

$$\min_{(k_P,k_R)\in C} \sum_{j=1}^{n} \|\bar{x}_j - x_{\bar{u}}(t_j)\|^2 \quad \text{s.t.} \quad \left| \begin{array}{rl} \frac{dc}{dt} = & k_P p - k_R c r - k_P p c, \\ \frac{dp}{dt} = & u k_R c r - k_P p^2, \\ \frac{dr}{dt} = & (1-u) k_R c r - k_P p r, \\ \bar{u}(t) \in & \arg\max_{u(\cdot)} \int_0^T k_P p(t)\, dt, \\ & 0 \le u(t) \le 1 \quad \text{a.e.}\, t \in [0,T], \end{array} \right. \tag{1}$$

(2)

in which c, p, and r represent respectively the proportion of carbon reserve (precursors), photosynthetic machinery (e.g., the light-harvesting complex LHCII), and gene expression machinery (e.g., ribosomal proteins), $x := (c,p,r)$, and $u(\cdot)$ is the allocation of protein (the control variable). Experimental measurements yield to the values of $\bar{x}_j$. The set $C \subset \mathbb{R}^2$ is compact and is the admissible set for parameters of the model.
We shall first determine using Pontryagins' Principle and numerical simulations the optimal strategy, corresponding to a turnpike with a chattering arc. Then, a bi-level optimization problem (as above) is proposed to calibrate the model with experimental data. To solve it, a classical parameter identification routine is used, calling at each iteration the bocop solver to solve the optimal control problem (by a direct method). The calibrated model is able to represent the photoacclimation dynamics of the microalga *Dunaliella tertiolecta* facing a down-shift of light intensity.

References

[1] I. Yegorov, F. Mairet, J.-L. Gouzé, *Optimal feedback strategies for bacterial growth with degradation, recycling, and effect of temperature*, Optimal Control Applications and Methods, 2 (2018), pp. 1084–1109.

On the Interplay between Acceleration and Identification for the Proximal Gradient algorithm

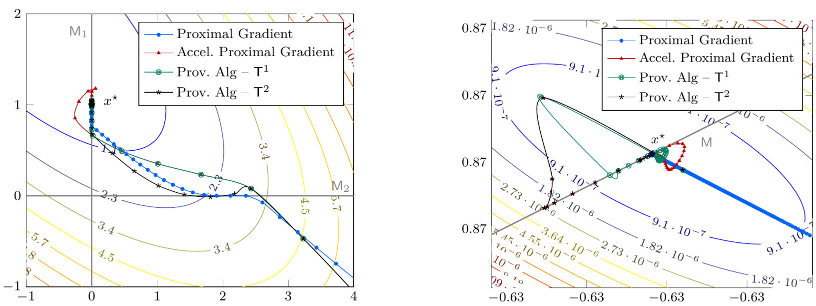
Gilles Bareilles
Univ. Grenoble Alpes

Keywords : Proximal Gradient, Accelerated methods, Identification.

Many problems in machine learning or linear inverse problems are formulated as composite non-smooth optimization problems. The non-smoothness is chosen to enforce some desirable structure on the solution, but it also affects the iterates produced by optimization methods. In this talk, we discuss the conditions under which the Proximal Gradient algorithm and its accelerated counterpart [1, 2] are able to collect this structure and how they do so.

Identification We are interested in solving $\min_{x\in\mathbb{R}^n} f(x) + g(x)$, where f is a differentiable convex function and g is a convex function, generally non differentiable.
As previously mentioned, the non-differentiable function g is chosen to enforce some structure on the optimal solution. For example, coordinate sparsity is enforced by taking g as the ℓ_1-norm, and yields more interpretable models for linear regression (in which case, the problems boils down to the lasso). Besides, the iterates of a given optimization method progressively collect that optimal structure, in a more or less stable way. When they reach (and then stay in) that optimal structure, we say that *identification* happens.

(Accelerated) Proximal Gradient Besides ensuring identification, it is desirable that iterates collect the optimal structure in a *stable* manner. Indeed, a stable identification behavior means that stopping the algorithm before convergence still provides some relevant structural information. Based on [3], we show that the Proximal Gradient does identify in a stable way, but its accelerated counter part has a more erratic behavior. We also propose two *Provisional algorithms* that benefit from the $\mathcal{O}(1/k^2)$ rate of accelerated methods while being more stable.



References

[1] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM journal on imaging sciences, vol. 2 (2009).

[2] Y. Nesterov, *A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$*, Dokl. akad. nauk Sssr, vol. 269 (1983).

[3] G. Bareilles and F. Iutzeler, *On the Interplay between Acceleration and Identification for the Proximal Gradient algorithm*, preprint arXiv:1909.08944 (2019).

Decomposition of a high dimensional stochastic optimization problem and application to flexible devices in the context of Smart Grids

SEGURET Adrien
FiME research center

Joint work with *C. ALASSEUR (EDF RD), F. BONNANS (CMAP), A. DE PAOLA (U. Bath), N. OUDJANE (EDF RD)* and *V. TROVATO (Imperial College London)*

Keywords : Stochastic Control, Decomposition, Power System, Smart Grids

In the power sector, many smart and flexible devices are currently appearing at the customer side. The issue on how to manage all those devices in an efficient way is an important and difficult question especially due to the high number of devices and the stochastic nature of the problem. Our work aims at providing a decomposition algorithm for solving this complex problem. We first consider the following stochastic optimization problem (P_1) involving n agents or devices, where X^{i,u^i} is a solution of a SDE.

$$(P_1) \quad \inf_{u\in\mathcal{U}} \mathbb{E}\big[F_0(\frac{1}{n}\sum_{i=1}^{n} u^i) + \sum_{i=1}^{n} F_i(u^i, X^{i,u^i})\big], \tag{1}$$

The price decomposition approach appears to be unfeasible since the price depends on the history of all agents. Therefore we consider a modified problem:

$$(P_2) \quad \inf_{u\in\mathcal{U}} F_0(\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} u^i]) + \sum_{i=1}^{n} \mathbb{E}\big[F_i(u^i, X^{i,u^i})\big] \tag{2}$$

Under some reasonable assumptions, the solutions of (P_2) are ε-optimal control of (P_1). In addition, (P_2) can be decomposed with a deterministic price (this can be compared to [1]). The resulting algorithm can be viewed as a randomized Uzawa algorithm, whose convergence is established.

We apply the proposed algorithm to the context of a power system motivated by the work of A. De Paola *et al.* [2]. In this problem, a central planner solves a Unit Commitment (UC) problem (i.e. the optimization problem for determining the on-line generation capacities, its production level, including security margins, to match the system load, which includes the coordination of the consumption....) while coordinating the operation of a large population of smart devices.

References

[1] Pierre Carpentier, Jean Philippe Chancelier, Vincent Leclère, and François Pacaud. Stochastic decomposition applied to large-scale hydro valleys management. *European Journal of Operational Research*, 270(3):1086–1098, 2018.

[2] Antonio De Paola, Vincenzo Trovato, David Angeli, and Goran Strbac. A mean field game approach for distributed control of thermostatic loads acting in simultaneous energy-frequency response markets. *IEEE Transactions on Smart Grid*, PP:1–1, 01 2019.

Hard Shape-Constrained Kernel Regression

Pierre-Cyril Aubin-Frankowski
École des Ponts ParisTech and CAS, MINES ParisTech, PSL Research University

Joint work with *Zoltán Szabó*

Keywords : Kernel methods, Shape constraints, Non-parametric regression

Shape-constrained regression problems [1, 2, 3] arise in a large number of applications. Economic theory dictates that utility functions are increasing and concave, demand functions of normal goods are downward sloping, production functions are concave or S-shaped, or that the link function in a single index model is typically monotone. In finance, European and American call option prices are convex and monotone in the underlying stock price and increasing in volatility, bond yield curves are monotone and concave in time to maturity, the conditional value-at-risk measure is increasing w.r.t. the significance level. In stochastic control and reinforcement learning the value function is regularly assumed to be convex.

Leveraging prior knowledge expressed in terms of shape structures has several advantages: the resulting techniques allow for estimation with smaller sample size, handle larger scale tasks, and help interpretability. Despite the numerous practical benefits the construction of shape-constrained estimators is quite challenging, existing techniques often tackle the shape requirements (i) in a soft fashion (without out-of-sample guarantees), (ii) by specialized transformation of the variables (such as logarithmic or translog specifications), or (iii) using of highly restricted functions classes such as polynomial splines.

In this work, we focus on the problem of shape-constrained regression with pointwise inequality constraints:

$$\min_{f\in\mathcal{F}_k} L\left((\mathbf{x}_n, y_n, f(\mathbf{x}_n))_{n\in[\![1,N]\!]}\right) + R\left(\|f\|_k\right), \text{ subject to } b_0 \leq D(f-f_0)(\mathbf{x})\ \mathbf{x}\in K_0, \tag{1}$$

where the hypothesis space is a reproducing kernel Hilbert space $\mathcal{F}_k$ of $\mathbb{R}^d \to \mathbb{R}$ functions; the samples $\{(\mathbf{x}_n, y_n)\}_{n\in[\![1,N]\!]}$, the regularization function R, the differential operator D, the constant b_0, the function f_0, and the compact nondiscrete set K_0 are given.

We show how second-order cone programming techniques can be applied to solve a strengthened version of (1), and hence to satisfy *strictly* the imposed shape constraints. In addition, we provide performance guarantees w.r.t. to the solution of the original problem (1). We demonstrate the efficiency of the proposed technique in joint quantile regression and in the context of transportation systems.

References

[1] D. Chetverikov, A. Santos and A. M. Shaikh, *The Econometrics of Shape Restrictions*, Annual Review of Economics, 10, (2018), pp. 31-63.

[2] A. L. Johnson and D. R. Jiang, *Shape Constraints in Economics and Operations Research*, Statist. Sci., 33 (4), (2018), pp. 527-546.

[3] Adityanand Guntuboyina and Bodhisattva Sen, *Nonparametric Shape-Restricted Regression*, Statist. Sci., 33 (4), (2018), pp. 568-594.

Online and Stochastic Optimization beyond Lipschitz continuity: A Riemannian Aproach

Antonakopoulos Kimon
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP

Joint work with *E.V. Belmega, P. Mertikopoulos*

Keywords : Online convex optimization, Regret analysis, Stochastic optimization, Non-Convex optimization

First-order methods are still one of the cornerstones of optimization. One of the main reasons for this is that the computation of higher-order derivatives of functions with thousands if not millions of variables quickly becomes prohibitive; another is that gradient calculations are typically easier to distribute and parallelize, especially in large-scale problems. In view of this, first-order methods have met with prolific success in many diverse fields, from machine learning and signal processing to wireless communications, nuclear medicine etc. This success is especially pronounced in the field of *online optimization*, i.e. when the optimizer faces a sequence of time-varying loss functions $f_t(x)$, $t = 1, 2, \ldots$, one at a time for instance, when drawing different sample points from a large training set. In this general framework, first-order methods have proven extremely flexible and robust, and the attained performance guarantees are well known to be optimal. Specifically, if the optimizer faces a sequence of G-Lipschitz convex losses, the incurred min-max regret after T rounds is $\Omega(GT^{1/2})$ [1] , and this bound can be achieved by inexpensive first-order methods such as (OMD) and its variants.
Nevertheless, in many machine learning problems (support vector machines, Poisson inverse problems, quantum tomography, etc.), the loss landscape is *not* Lipschitz continuous, so the results mentioned above do not apply. Thus, a natural question that emerges is the following: Is it possible to apply online optimization tools and techniques beyond the standard Lipschitz framework? And, if so, how? The answer would positive and is illustrated in our paper [2]. In particular, we introduce and examine a (RL) continuity condition which is tailored to the singularity landscape of the problem's loss functions. In this way, we are able to tackle cases beyond the Lipschitz framework provided by a global norm, and recover the optimal regret bounds that are known for the standard case. These results are applied for the stochastic convex minimization framework.
Secondly , we show a last iterate convergence result for the stochastic Mirror Descent, (i.e. establishing the convergence of the actual sequence generated by the algorithm). This is of particular interest for non-convex problems where ergodic convergence results are of limited value (because Jensen's inequality no longer applies). These results are subsequently validated in a class of stochastic Poisson Inverse Problems that arise in imaging science.

References

[1] Martin Zinkevich, *Online Convex programming and generalized infinitesimal gradient ascent*, ICML '03: Proceedings of the 20th International Conference on Machine Learning. 928–936.

[2] K.Antonakopoulos, E. V. Belmega, P.Mertikopoulos, *Online and stochastic beyond Lipschitz continuity:A Riemannian approach*, in ICLR 2020, forthcoming.

Performance Analysis of Chaos Optimization Algorithm Based on Quadratic maps

Hamaizia Tayeb
Department of Mathematics, Faculty of Sciences Exacts, University Constantine 1 Alger
h2tayeb@gmail.com
Djafri Nabila ***nabiladj540@gmail.com***

Keywords : Global Optimization,chaotic optimization, chaos

1 Abstract:

Chaos optimization algorithm is an effective global optimization algorithm with low computational complexity and fast search speed. The ergodicity of chaos is a new and interesting property for optimization, of which the main idea is to search according to a series of successive fields; first search in the total field, then in a smaller one, and so on. The most famous phenomena of chaos are described by Logistic equation. But the distribution of chaotic sequences produced by logistic map is non-uniform leading to the slow constringency. The Quadratic map is uniform map with which we replace logistic map to accelerate the rate of convergence. The conclusion is confirmed by simulation. Therefore it is of great importance to applying Quadratic equation in global optimization

References

[1] R. Caponetto, L. Fortuna, S. Fazzino and M. G. Xibilia, Chaotic Sequences to Improve the Performance of Evolutionary Algorithms, IEEE Transactions on Evolutionary Computation, Vol.7,n3,June2003,289-304.

[2] T. Hamaizia, R. Lozi, An improved chaotic optimization algorithm using a new global locally averaged strategy, Journal of Nonlinear Systems and Applications 3 (2) (2012) 58-63.

[3] L. S. Coelho, Tuning of PID controller for an automatic regulator voltage system using chaotic optimization approach, Chaos, Solitons and Fractals, 39, 2009, 1504-1514.

Sequential network formation: new stability concepts

Julien Fixary
Université Paris 1 Panthéon-Sorbonne

Joint work with *Philippe Bich*

Keywords : Pairwise stability, subgame perfect equilibrium, sequential network formation

Abstract

For twenty years, literature on networks has been very active. Part of this literature models network formation using a game theoretic approach: each node in the network is an agent and every agent is given a payoff function (a representation of its linking preferences) and chooses to form links with other agents based upon it. A standard modelization is to consider the notion of *society* which can be seen as the counterpart of a normal-form game in network formation and which is composed of a *set of agents* N (a finite subset of $\mathbb{N}$), a *set of weights* W (a subset of $[0,1]$), a *set of links* L (a subset of $\{\{i,j\} \subset N \mid i \neq j\}$, where $\{i,j\} \in L$ is denoted ij) and for every agent $i \in N$, a *payoff function* v_i (a mapping from the *set of networks* $\mathbb{G} := W^L$ to $\mathbb{R}$). In this context, Jackson and Wolinsky introduced a strategic notion of stability which is called *pairwise stability*. Briefly, a network $g \in \mathbb{G}$ is pairwise stable if for every link ij, (1) there exists no agent $k \in ij$ which has a strict interest (regarding v_k) to decrease the weight $g(ij)$ attributed to ij, (2) both agents of i and j don't have a common strict interest (regarding v_i and v_j) to increase the weight $g(ij)$ attributed to ij. The purpose of our work is to extend the idea of this concept to a sequential framework.

First, we introduce a possible counterpart of an extensive-form game in network formation that we call *sequential society*. Intuitively, we consider a network formation process in T steps ($T \in \mathbb{N}^*$) such that at each step t corresponds a link $i_t j_t$. Agents i_t and j_t have to choose a weight w_t regarding every possible choices of every links $i_\tau j_\tau$ ($\tau < t$).

Second, given a sequential society, we introduce two possible counterparts of the concept of subgame perfect equilibrium in network formation that we call *sequentially pairwise stable equilibrium* (or *SPSE*) and *weak sequentially pairwise stable equilibrium* (or *weak SPSE*). These two concepts share some features with the concept of subgame perfect equilibrium and with the concept of pairwise stability introduced by Jackson and Wolinsky. By definition of pairwise stability, it can be expected that finding a SPSE (or a weak SPSE) leads to an unusual optimization problem. We study these notions and the existence of elements satisfying its in three different cases (depending on the set of weights W): when W corresponds to $\{0,1\}$ (*unweighted case*), when W is a finite subset of $[0,1]$ (*discretely weighted case*) and when W corresponds to $[0,1]$ (*weighted case*).

References

[1] M.O. Jackson and A. Wolinsky, *A strategic model of social and economic networks*, Journal of Economic Theory (1996).

[2] P. Bich and L. Morhaim, *On the existence of pairwise stable weighted networks*, Mathematics of Operations Research (to appear), HAL-SHS (2017 - preprint).

A Decomposition Method by Component for the Optimization of Maintenance Scheduling

T. Bittar
EDF R&D, PRISME Department / CERMICS, École des Ponts ParisTech

Joint work with *P. Carpentier, J-Ph. Chancelier,* and *J. Lonchampt*

Keywords : Maintenance scheduling, decomposition-coordination, stochastic optimization

In order to improve the reliability and performance of its hydroelectric fleet, EDF (Électricité De France) seeks to optimize the maintenance scheduling of the components of hydroelectric power plants. In this work, we address an idealized maintenance scheduling optimization problem. We consider components of hydroelectric power plants such as turbines, transformers or generators and we study an idealized system of a given type of components that share a common stock of spare parts. Over time components experience random failures that occur according to known failure distributions. Thus the dynamics of the system is stochastic. The goal is to find a deterministic preventive maintenance strategy that minimizes the expected cost depending on maintenance and on the occurrences of forced outages of the system. The numerical experiments should involve systems constituted of up to 80 components thus leading to high-dimensional optimization problems. To overcome the curse of dimensionality, we use the Interaction Prediction Principle [1] to decompose the original optimization problem into a sequence of independent subproblems of smaller dimension. Each subproblem consists in optimizing the maintenance on a single component. The resulting algorithm iteratively solves the subproblems with the blackbox algorithm MADS [3] and coordinates the components. The maintenance optimization problem is a mixed-integer problem. However decomposition methods are based on variational techniques so they are not suited for this kind of problems. It is therefore needed to relax the dynamics of the system as well as the cost functions in the formulation of the maintenance optimization problem. The relaxation parameters have an important influence on the output of the optimization by decomposition and must be appropriately chosen. We apply the decomposition method on relaxed systems with up to 80 components. The more demanding case takes around 20 hours of computation time. We show that in high dimension the decomposition outperforms the blackbox algorithm applied directly on the original problem.

References

[1] M. D. Mesarović, D. Macko and Y. Takahara, *Theory of Hierarchical, Multilevel Systems*, volume 68 of *Mathematics in Science and Engineering*, Academic Press, 1970.

[2] G. Cohen, Auxiliary problem principle and decomposition of optimization problems. *Journal of Optimization Theory and Applications*, 32(3):277-305, 1980.

[3] C. Audet, J. E. Dennis Jr., Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization*, 17(1):188-217, 2006.

About optimal crop irrigation planing under water scarcity

Kenza Boumaza*

Joint work with *A. Rapaport and *S. Roux***

*UMR MISTEA, Univ. Montpellier, INRA-SupAgro Montpellier

Keywords : Optimal Control, state constraint, water management

Irrigation scheduling is the process of determining the amount of water delivered using an irrigation system during the production season. Several methodological approaches have been used with the aim of defining optimal strategies for irrigation scheduling. However, most approaches lack analytical insight on the theoretical properties of optimal solutions. In this work, we consider a simplified dynamical model of crop irrigation, inspired from [2], where $S(t)$ and $B(t)$ stand respectively for the relative soil humidity (a number between 0 and 1) and the crop biomass at time t belonging to an interval $[0, T]$ representing the crop growth season:

$$\begin{aligned} \dot{S} &= k_1(-\varphi(t)K_S(S) - (1-\varphi(t))K_R(S) + k_2 u(t)) \quad (1)\\ \dot{B} &= k_3\varphi(t)K_S(S) \quad (2) \end{aligned}$$

with the initial condition (at the sowing date 0)

$$S(0) = 1 \quad \text{and} \quad B(0) = 0 \quad (3)$$

and T being the harvesting date. The control variable $u(t) = F(t)/F_{max} \in [0, 1]$ is the ratio of the input water flow rate $F(t)$ at time t over the maximal flow F_{max} that the irrigation device allows. $K_S(\cdot)$ and $K_R(\cdot)$ are continuous piece-wise linear functions from $[0, 1]$ to $[0, 1]$. They represent the transpiration and the evaporation of the crop cultures respectively. $\varphi(t)$ representing the crop radiation use efficiency and independent of water stress, is a L^1 increasing function from $[0, T]$ to $[0, 1]$ with $\varphi(0) = 0$ and $\varphi(T) = 1$. And k_1, k_2, k_3 are positive parameters. The objective in this work, in the spirit of [3], is to study admissible strategies $u(\cdot)$ maximizing the biomass production $B_T[u(\cdot)]$ under a constraint of a given amount of water available for the period

$$Q[u(\cdot)] \leq \bar{Q}. \quad (4)$$

For this particular problem we give a sufficient condition on the target that ensures that any optimal solution does not saturate the state constraint. Despite the lack of differentiability of the functions $K_R(\cdot)$ and $K_S(\cdot)$ we show, by using the generalized maximum principle [1], that the optimal solution may present a singular arc under a lower water availability. Which leads to a bang-singular-bang solution. This solution may be better than a simple bang-bang control as commonly used.

References

[1] F. H. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer-Verlag, London, (2013).

[2] N. Pelak, R. Revellia and A. Porporato, *A dynamical systems framework for crop models: Toward optimal fertilization and irrigation strategies under climatic variability*, Journal of Ecological Modelling, vol. 365, (2017), pp. 80-92.

[3] U. Shani, Y. Tsur and A. Zemel, *Optimal dynamic irrigation schemes*, Journal of Optimal Control Applications and Methods, vol. 25, (2004), pp. 91-106.

Spatial DTW: Optimal transport in space and time

Hicham Janati
Inria Saclay / ENSAE

Joint work with *Marco Cuturi* and *Alexandre Gramfort*

Keywords : Optimal transport, Time series, Clustering

This talk will introduce Spatio-temporal alignments (STA). A new metric for time series that (1) takes into account their chronological order; (2) quantifies temporal shifts; (3) is sensitive to spatial variability. For instance, In neuroscience, we would like to compare brain imaging signals both in space and time. Let $\mathbf{x}_1, \mathbf{y}_2 \in \mathbb{R}^{T_1,p} \times \mathbb{R}^{T_2,p}$ be two p dimensional time series of lengths T_1, T_2 respectively. Let $(\Delta_{t_1,t_2}) \in \mathbb{R}^{T_1,T_2}$ be a pairwise cost matrix between all time points of $\mathbf{x}, \mathbf{y}$. A feasible alignment is defined as a binary matrices of $\mathbb{R}^{T_1 \times T_2}$ where only $\rightarrow, \downarrow, \searrow$ movements are allowed. This set is denoted by $\mathcal{A}_{T_1,T_2}$. We build upon the Soft Dynamic Time Warping (DTW) framework [2, 3] defined as: $\mathbf{dtw}_\beta(\mathbf{x}, \mathbf{y}; \Delta) = \mathrm{softmin}_\beta\{\langle \mathbf{A}, \Delta(\mathbf{x}, \mathbf{y})\rangle, \mathbf{A} \in \mathcal{A}_{T_1,T_2}\}$where the soft-minimum operator of a set $\mathcal{A}$ with parameter β is defined as $-\beta \log\left(\sum_{\mathcal{A}} e^{-a/\beta}\right)$ if $\beta > 0$ and $\min_{a \in \mathcal{A}} a$ if $\beta = 0$. Here we propose a cost Δ defined with unbalanced Optimal transport.

1 Temporal OT through Soft-DTW

Our first contribution is to show that when $\beta > 0$, $\mathbf{dtw}_\beta$ increases quadratically with time shifts:

Theorem 1. *Let $\mathbf{x}_+$ correspond to the shifted $\mathbf{x}$ with lag k. For any cost Δ, let μ the smallest non-zero entry of Δ. If $0 < \beta \leq \frac{\mu}{\log(3TD_{T,T})}$, there exists $\alpha, \rho > 0$ such that:*

$$\mathbf{dtw}_\beta(\mathbf{x}, \mathbf{x}_{+_k}) - \mathbf{dtw}_\beta(\mathbf{x}, \mathbf{x}) \geq \beta\alpha k(k-1) + \beta\rho k \tag{1}$$

2 Spatial OT through Sinkhorn divergences

To capture spatial variability, we propose to use a cost function based on the unbalanced Sinkhorn of cost M distance OT_ε[1] defined for any non-negative histograms in $\mathbb{R}^p$: $\Delta_{\mathrm{OT}}(\mathbf{x}, \mathbf{y}) = \mathrm{OT}_\varepsilon(\mathbf{x}, \mathbf{y}) - \frac{1}{2}\left(\mathrm{OT}_\varepsilon(\mathbf{x}, \mathbf{x}) + \mathrm{OT}_\varepsilon(\mathbf{y}, \mathbf{y})\right)$ where $\mathrm{OT}_\varepsilon(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{P} \in \mathbb{R}_+{}^{p \times p}} \varepsilon \mathrm{KL}(\mathbf{P}|e^{-\frac{\mathbf{M}}{\varepsilon}}) + \gamma \mathrm{KL}(\mathbf{P}\mathbb{1}|\mathbf{x}) + \gamma \mathrm{KL}(\mathbf{P}^\top \mathbb{1}|\mathbf{y})$. We simulate 200 brain signals on a spatial mesh of $p = 642$ vertices and $T = 20$ in 4 cross-categories in space and time and learn t-SNE embeddings with different metrics.

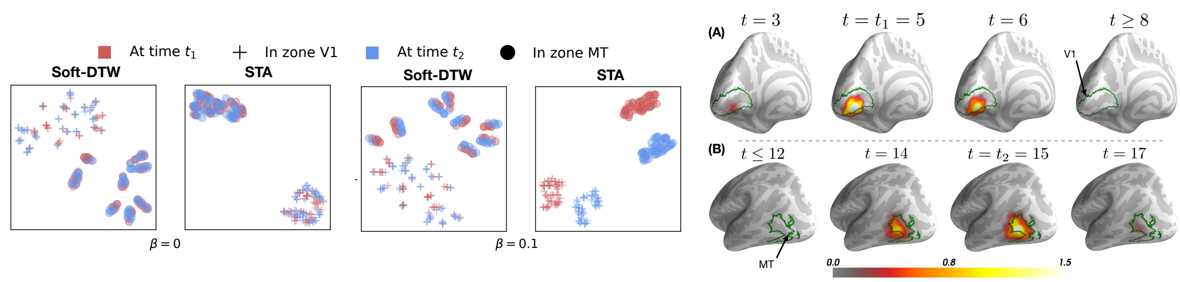


Figure 1: t-SNE visualization of the simulated brain signals in two different regions, at two different time instants. With $\beta > 0$, $\mathbf{sta}_\beta$ can discriminate between all four groups.

References

[1] L. Chizat, G. Peyré, B. Schmitzer, and F-X. Vialard. Scaling Algorithms for Unbalanced Transport Problems. *arXiv:1607.05816 [math.OC]*, 2017.

[2] Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *International Conference on Machine Learning*, 2017.

[3] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, February 1978.

Two-player zero-sum deterministic differential game in infinite horizon involving impulse controls with a new QVI

HAFID LALIOUI
Université Ibn Zohr, Equipe, Aide à la decision, ENSA, Agadir, Morocco

Joint work with *BRAHIM EL ASRI*

Keywords : Deterministic differential game, impulse controls, dynamic programming principle, viscosity solutions, quasi-variational inequality.

We study a two-player zero-sum deterministic differential game with both players adopting impulse controls in infinite horizon with costs functions depends on the state variable. We aim to prove that the impulse control problem admit a value.
We prove by the mean of dynamic programming principle that the value functions are continuous and viscosity solutions of the corresponding Hamilton-Jacobi-Bellman-Isaacs (HJBI) quasi-variational inequality (QVI). We define a new differential QVI for which the value functions are the unique viscosity solution. We then prove that the lower and upper value functions coincide.
In this work we deal with the state variable $y_x(.)$ such that $y_x(0) = x \in R^n$, $\dot{y}_x(t) = b(y_x(t))$ and $y_x(.)$ is driven by two impulse controls, u control of $player - \xi$ defined by a double sequence $(\tau_m, \xi_m)_{m \in N^*}$ and v control of $player - \eta$ defined by a double sequence $(\rho_k, \eta_k)_{k \in N^*}$.
We give some classic results of the lower and upper value functions of our game, first we show that both satisfy the dynamic programming principle property, then we prove that they are continuous in R^n.
The associated QVI is given by a double obstacles HJBI equation, where the obstacles are defined in function of the infimum (resp. maximum) cost operator $\mathcal{H}^{\chi}_{inf}v(.)$ (resp. $\mathcal{H}^{c}_{sup}v(.)$). We prove the existence of the value functions as a viscosity solution of this classic HJBI QVI.
Furthermore, we consider a new differential QVI for which the existence and the uniqueness result in the viscosity solution sense is guaranty. Thus, the game admits a value.

References

[1] N. El Farouq, G. Barles and P. Bernhard, *Deterministic minimax impulse control*, Appl. Math. Optim., 61 (2010), pp. 353–378, DOI: 10.1007/s00245-009-9090-0.

[2] B. El Asri, *Deterministic minimax impulse control in finite horizon: the viscosity solution approach*, ESAIM: Control Optim. Calc. Var., 19 (2013), pp. 63–77, DOI: 10.1051/cocv/2011200.

[3] G. Barles, *Deterministic impulse control problems*, SIAM J. Control Optim., 23 (1985), pp. 419–432.

[4] N. El Farouq, *Degenerate first-order quasi-variational inequalities: an approach to approximate the value function*, SIAM J. Control Optim., 55 (2017), pp. 2714–2733.

[5] H. Ishii, *Perron's method for Hamilton-Jacobi equations*, Duke Math. J., 55 (1987), pp. 369–384.

[6] P.L. Lions, *Generalized Solution of Hamilton-Jacobi Equations*, Pitman, Boston (1982).

Stable Matching Games

Felipe Garrido-Lucero
Université Paris Dauphine - PSL

Joint work with *Rida Laraki*

Keywords : Stable Matching, Nash Equilibrium with Outside Option Constraints

Gale and Shapley [1] defined in 1962 a matching problem between two sets of distinct agents M and W (men and women, workers and firms, students and universities, nurses and hospitals, kidneys and patients, etc) who should be paired in a *stable* way, taking into account that each agent has a strict order of preference over those members of the opposite set. A matching μ is stable if no unmatched couple can strictly improve its utility by being together compared to their utility in μ. They constructed, thanks to an algorithm, a stable matching for every instance. In 1971, Shapley and Shubik [3] considered an extension in which agents can do monetary transfers that can change their quasi-linear utility. A solution to this new problem is a pair (μ, p) with μ a matching between buyers and sellers and p a vector of monetary transfers. A solution is stable if no pair of unpaired agents can find a transfer which improves their utility compared with (μ, p). In 2005, Milgrom and Hatfield [2] extended this model to *Matching with contracts* in which agents define partners and bilateral contracts, from a set of possible options. They provide sufficient conditions on the set of contracts under which a stable allocation exists.
Our article proposes a new extension by assuming that each couple $(i, j) \in M \times W$ obtains its utility as the outcome of a strategic game $G_{i,j}$. In this approach, agents, beside choosing partners, play a strategic game in which they try to maximize their outcome under the constraint that the partner is satisfied and does not want to quit the relationship. A matching μ is externally stable if no unmatched couple can deviate, i.e. to match together and play a strategy profile which provides them higher payoffs than the ones they have in μ. This corresponds to the classical pairwise stability notion. The matching is internally stable if no player can, by changing her strategy inside the couple, increases her payoff without breaking the external stability of the couple. Mathematically, internal stability is simply a Nash equilibrium condition with outside option constraints. We provide a sufficient condition on a strategic game (called feasibility) under which there exists matching profiles that are externally and internally stable. We prove that the class of *feasible games* includes zero-sum games, potential games, infinitely repeated games or finite perfect information games. We also provide a 3 by 3 game in mixed strategies which is not feasible. Since monetary transfers can be interpreted as zero sum games, our model strictly includes Shapley-Shubik's extension. Similarly, we also extend Milgrom-Hatfield's model because any stable allocation can be mapped into a externally stable matching and conversely.

References

[1] David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[2] John William Hatfield and Paul R Milgrom. Matching with contracts. *American Economic Review*, 95(4):913–935, 2005.

[3] Lloyd S Shapley and Martin Shubik. The assignment game i: The core. *International Journal of game theory*, 1(1):111–130, 1971.

Sinkhorn Divergences for Unbalanced Optimal Transport

Thibault Séjourné
DMA, ENS, France

Joint work with *Jean Feydy, F-X Vialard, Alain Trouvé, Gabriel Peyré*

Keywords : Optimal transport, Entropic regularization, Sinkhorn divergence, Unbalanced OT

We will introduce in this presentation an extension the formulation of Sinkhorn divergences to the unbalanced setting of arbitrary positive measures, providing both theoretical and algorithmic advances. Sinkhorn divergences, introduced in [1], leverage the entropic regularization of Optimal Transport (OT) to define geometric loss functions. They are differentiable, cheap to compute and do not suffer from the curse of dimensionality, while maintaining the geometric properties of OT, in particular they metrize the weak* convergence. Extending these divergences to the unbalanced setting is of utmost importance since most applications in data sciences require to handle both transportation and creation/destruction of mass. This includes for instance problems as diverse as shape registration in medical imaging, density fitting in statistics, generative modeling in machine learning, and particles flows involving birth/death dynamics. Our first set of contributions is the definition and the theoretical analysis of the unbalanced Sinkhorn divergences. They enjoy the same properties as the balanced divergences (classical OT), which are obtained as a special case. Indeed, we show that they are convex, differentiable and metrize the weak* convergence. Our second set of contributions studies generalized Sinkhorn iterations, which enable a fast, stable and massively parallelizable algorithm to compute these divergences. We show, under mild assumptions, a linear rate of convergence, independent of the number of samples, i.e. which can cope with arbitrary input measures. We also highlight the versatility of this method, which takes benefit from the latest advances in term of GPU computing, for instance through the KeOps library for fast and scalable kernel operations.

References

[1] FEYDY, Jean, SÉJOURNÉ, Thibault, VIALARD, François-Xavier, et al. Interpolating between optimal transport and MMD using Sinkhorn divergences. arXiv preprint arXiv:1810.08278, 2018.

[2] CHIZAT, Lenaic, PEYRÉ, Gabriel, SCHMITZER, Bernhard, et al. Scaling algorithms for unbalanced optimal transport problems. Mathematics of Computation, 2018, vol. 87, no 314, p. 2563-2609.

[3] LIERO, Matthias, MIELKE, Alexander, et SAVARÉ, Giuseppe. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. Inventiones mathematicae, 2018, vol. 211, no 3, p. 969-1117.

Diffusion in large networks

Michel Grabisch
Université Paris I Panthéon-Sorbonne, Paris School of Economics

Joint work with *Agnieszka Rusinowska* and *Xavier Venel*

Keywords : diffusion, networks, opinion dynamics

Let $\mathcal{X}$ be the society of agents, supposed to be countably infinite. A neighborhood relation is defined on $\mathcal{X}$, which is symmetric and irreflexive, inducing a graph structure on $\mathcal{X}$ which is connected and where each agent has a finite degree. At a given time, each individual has two possible statuses (either active (1) or inactive (0)). The status of an agent is updated by aggregating the statuses of the agents in its neighborhood. The number obtained is the probability to be active at next time step for that agent. Defining the state (or configuration) of the society as the set of active agents, this defines an uncountable Markov chain in discrete time. The aim of the paper is to study the absorbing classes of this Markov chain. We distinguish the case where the aggregation function is strict or Boolean.
The main part of the study is devoted to the strict case. It is found that:

- Under no additional assumption on the graph nor on the aggregation function, the sets of all finite and all co-finite configurations are transient, the sets $\mathcal{X}$ and $\emptyset$ are fixed points (absorbing states), while the set of all remaining configurations is absorbing. In addition, if the graph admits a bipartite structure, periodic classes of period 2 occur (called blinkers).
- The establishment of irreducibility for the previous sets of configuration requires however stronger assumptions on the graph. We proved that sufficient conditions are:
  1. There are infinitely many complex stars (roughly speaking, stars with at least 3 branches of length at least 2)
  2. The network has enough "space" to store a local configuration of active/inactive agents on its neighbors.

  Most networks satisfy this assumption and we characterize networks having complex stars.

Under Boolean aggregation functions, the diffusion process becomes deterministic and the contagion model of [1] becomes a particular case of our framework. Then finite and co-finite absorbing states can exist, as well as cycles.

References

[1] S. Morris, *Contagion*, *Review of Economic Studies*, 67:57–78, 2000.

Maximisation de déterminants de Sturm-Liouville

Jean-Baptiste Caillau
Université Côte d'Azur, CNRS, Inria, LJAD

Mots-clés. Déterminant fonctionnel, fonction ζ spectrale, opérateur de Schrödinger

On recherche des potentiels qui maximisent le déterminant fonctionnel d'un opérateur de Sturm-Liouville du type Schrödinger sur un intervalle borné, avec conditions aux limites de Dirichlet. L'optimisation est faite sous contrainte L^q. On étend pour cela la définition du déterminant fonctionnel au cas de potentiels L^1, et on montre que le problème de maximisation est équivalent à un problème de contrôle optimal. On prouve l'existence et l'unicité de solution pour tout $q \geq 1$, et on donne une caractérisation des potentiels maximisants dans les cas $q = 1, 2$.

Travail en collaboration avec C. Aldana (Barranquilla) et P. Freitas (Lisbonne).

References

[1] Burghelea, D.; Friedlander, L.; Kappeler, T. On the determinant of elliptic boundary value problems on a line segment, *Proc. Amer. Math. Soc.* **123** (1995), 3027–3038.

[2] Harrell, E. M. Hamiltonian operators with maximal eigenvalues, *J. Math. Phys.* **25** (1984), 48–51; Erratum, J. Math. Phys. **27** (1986), 419.

[3] Levit, S.; Smilansky, U. A theorem of infnite products of eigenvalues of Sturm-Liouville type operators, *Proc. Amer. Math. Soc.* **65** (1977), 299–302.

[4] Ray, D. B.; Singer, I. M. R-torsion and the Laplacian on Riemannian manifolds, *Adv. Math.* **7** (1971), 145–210.

A DC approach for chance constrained problems

Yassine LAGUEL
Université Grenoble Alpes

Keywords : Stochastic programming, Chance constraints, Bilevel Programming

Chance constraints are a valuable tool for the design of safe and robust decisions in uncertain environments. In this talk, we propose an exact reformulation of a general chance constrained problem as a convex bilevel program. We then derive a tractable penalty approach for such problems. We finally release an open-source python toolbox with fast computational procedures.

Chance constrained problems Given two functions $f : \mathbb{R}^d \to \mathbb{R}$, $g : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, a random vector $\xi : \Omega \to \mathbb{R}^m$, and a safety probability level $p \in [0, 1)$, we consider the problem:

$$\begin{cases} \min_{x\in\mathbb{R}^d} & f(x) \\ \text{s.t.} & \mathbb{P}[g(x,\xi) \leq 0] \geq p \end{cases} \tag{1}$$

Intuitively, the constraint means that the inequality $g(x, \xi) \leq 0$ must be satisfied with high enough probability. However, because of the possible non-convexity it induces, even simple questions such as the convexity of the feasible set for the probability constraint can be hard [1]. A great part of the results produced so far on chance constrained problems rely on a parametric modeling of the chance constraint [2].
Our approach [3] consists in leveraging the link between quantiles and superquantiles [4] to derive an exact reformulation of (1) as a bilevel program where the lower problem is convex as soon as g is convex.

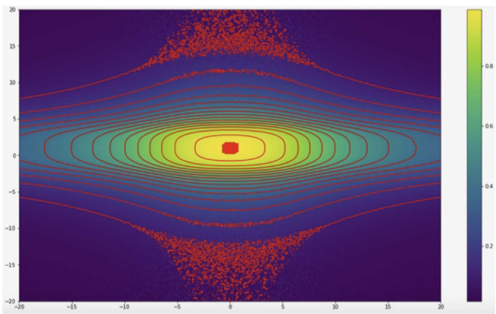


Figure 1: Levet sets of a quadratic chance constrained set [5]

Solving chance constrained problems with a difference-of-convex approach We propose to tackle the obtained bilevel problem with a double penalty approach. The penalized objective appears to be a Difference of Convex (DC) function that we propose to minimize with the bundle algorithm of [6]. We provide `TACO`, an open-source python **T**oolbox for ch**A**nce **C**onstrained **O**ptimization, for solving (1). The code is available at:

`https://github.com/yassine-laguel/taco`

References

[1] W. Van Ackooij, "Eventual convexity of chance constrained feasible sets," *Optimization*, 2015.
[2] R. Henrion and C. Strugarek, "Convexity of chance constraints with independent random variables," *Computational Optimization and Applications*, 2008.
[3] Y. Laguel, W. Van Ackooij, and J. Malick, "A data-driven approach for chance constrained problems," *In preparation.*
[4] R. T. Rockafellar and J. O. Royset, "Superquantiles and their applications to risk, random variables, and regression," in *Theory Driven by Influential Applications*, INFORMS, 2013.
[5] W. Van Ackooij, Y. Laguel, J. Malick, G. Matiussi Ramalho, and W. De Oliveira, "On the convexity of level-sets to probability functions," *Submitted.*
[6] W. de Oliveira, "Proximal bundle methods for nonsmooth dc programming," *Journal of Global Optimization*, 2019.

Hyperparameter selection for the Lasso with gradient descent

Quentin Bertrand
Université Paris Saclay, Inria, CEA

Joint work with *Q. Klopfenstein, J. Salmon, M.Blondel, S. Vaiter, A. Gramfort*

Keywords : Hyperparameter selection, Optimization, Non-smooth optimization, Sparsity

In many statistical applications, the number of parameters p is much larger than the number of observations n. A popular approach to tackle linear regression problems in such scenarios is to consider convex ℓ_1-type penalties, as popularized by [4]. The use of these penalties relies on a regularization parameter λ trading data fidelity versus sparsity, the latter being notoriously hard to tune in practice.
In order to prevent over complex models, it is customary to use different datasets for model training and hyperparameter selection. Classical approaches consist in evaluating each model with a *metric* on a *crossvalidation* dataset, different from the training dataset. Popular metrics includes *crossvalidation loss*, AIC/BIC [1] or SURE [3] criterions.
The canonical hyperparameter setting algorithm is the gridsearch, it requires evaluating the model on a given finite grid of parameters. Gridsearch is well adpated when the number of hyperparameters is small, however the complexity of the gridsearch scales exponentially with number of hyperparameters: *ie.* gridsearch cannot be applied when the number of hyperparameter is large. Other hyperparameter selection techniques include randomsearch and bayesian techniques (SMBO).
Another approach for hyperparameter setting consists in gradient descent in the hyperparameter to optimize. This has been widely explored in [2] for smooth objective function. In this work we tackle the setting of non-smooth hyperparameter optimization for Lasso type problems. We show that [2] techniques generalize to non-smooth optimization problems such as the Lasso [4] and the adaptive Lasso [5].
In gradient-based hyperparameter optimization techniques, the main challenge is to compute $\mathcal{J}_\lambda$: the weak jacobian of the function $\lambda \mapsto \hat{\beta}^{(\lambda)} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{2n} ||y - X\beta||^2 + \lambda ||\beta||_1$ at point λ. In this work propose an efficient implicit computation of $\mathcal{J}_\lambda$ for the Lasso and the adaptive Lasso.

References

[1] S. Liu and Y. Yang. Combining models in longitudinal data analysis. *Annals of the Institute of Statistical Mathematics*, 64:233–254, 2012.

[2] F. Pedregosa. Hyperparameter optimization with approximate gradient. In *ICML*, pages 737–746, 2016.

[3] C. M. Stein. Estimation of the mean of a multivariate distribution. In *Proc. Prague Symp. Asymptotic Statist.*, 1973.

[4] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 58(1):267–288, 1996.

[5] H. Zou. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476):1418–1429, 2006.

Prox-regular sets: separation and metric properties

Florent Nacry
University Perpignan Via Domitia

Joint work with *Samir Adly* and *Lionel Thibault*

Keywords : Variational analysis, prox-regular sets, separation, distance function.

It has been well recognized that *prox-regular sets* ([4]) (also know as *positively reached*, *weakly convex*, $\mathcal{O}(2)$*-convex*, φ*-convex*, *proximally smooth*) play an important role in many areas of mathematical analysis including Optimization, Set-Valued Analysis, Differential Geometry and PDES (see, e.g., the survey [2] and the references therein). Roughly speaking, a (closed) subset C of a Hilbert space X is said to be r*-prox-regular* provided that the nearest point mapping Proj_C is single-valued on a suitable neighborhood (depending on $r > 0$) of the set C and continuous therein.
In this talk, we present diverse new metric properties that prox-regular sets shared with convex ones. At the heart of our work lie the Legendre-Fenchel transform and complements of open balls. When the closed set C is r-prox-regular for some $r > 0$, it will be shown that the function

$$\varphi_{C,r}(x^\star) := (\psi_C + \frac{1}{2r}\|\cdot\|^2)^\star(x^\star) = \sup_{x\in C}(\langle x^\star, x\rangle - \frac{1}{2r}\|x\|^2)$$

is the right tool to extend the following fundamental results known for convex sets in convex analysis to the variational analysis of prox-regular sets:

(π_1) A closed convex set is completely determined by its support function $\sigma(\cdot,\cdot)$, i.e.,

$$C_1 = C_2 \Leftrightarrow \sigma(\cdot, C_1) = \sigma(\cdot, C_2).$$

(π_2) The analytic formulation of the distance from a convex set

$$d_C(x) = \langle x^\star, x\rangle - \sigma(x^\star; C) \quad \text{for some } x^* \in \mathbb{S} := \{u \in X : \|u\| = 1\}.$$

(π_3) The duality property

$$d_C(x) = \max_{x^\star\in\mathbb{S}} \inf_{y\in C} \langle x^\star, x - y\rangle = \inf_{y\in C} \max_{x^\star\in\mathbb{S}} \langle x^\star, x - y\rangle .$$

(π_4) The distance $d_C(x)$ coincides with the maximum of distances $d_H(x)$ taken over all hyperplanes H separating C and $x \notin C$ and this maximum is attained for one and only one hyperplane.

References

[1] S. Adly, F. Nacry, L. Thibault, *New metric properties for prox-regular sets*, submitted.

[2] G. Colombo, L. Thibault, *Prox-regular sets and applications*, Handbook of nonconvex analysis and applications 99-182 (2010), Int. Press, Somerville, MA.

[3] F.R. Deutsch, *Best Approximations in Inner Product Spaces*, CMS Books in Mathematics, Springer-Verlag, New-York, 2001.

[4] R.A. Poliquin, R.T. Rockafellar, L. Thibault, *Local differentiability of distance functions*, Trans. Amer. Math. Soc. 352 (2000), 5231-5249.

Constant Along Primal Rays Conjugacies Suitable for Functions of the Support

Michel De Lara
École des Ponts ParisTech

Joint work with *Jean-Philippe Chancelier*

Keywords : Optimization, sparse optimization, Fenchel-Moreau conjugacy

The support of a vector in $\mathbb{R}^d$ is the set of indices with nonzero entry. Functions of the support are widespread in sparse optimization, like the so-called ℓ_0 pseudonorm which counts the number of nonzero components of a vector. Functions of the support of a vector have the property to be 0-homogeneous. Because of that, the Fenchel conjugacy fails to provide relevant analysis. In this paper, we display a class of conjugacies that are suitable for functions of the support. For this purpose, we suppose given a (source) norm on $\mathbb{R}^d$. With this norm, we define, on the one hand, a family of so-called local coordinate-K norms (with $K \subset \{1, \ldots, d\}$) and, on the other hand, a coupling between $\mathbb{R}^d$ and $\mathbb{R}^d$, called Capra (constant along primal rays). Then, we provide formulas for the Capra-conjugate and biconjugate, and for the Capra-subdifferentials, of functions of the support in terms of the local coordinate-K norms. The special case of nondecreasing submodular functions of the support is considered.

References

[1] J.-P. Chancelier and M. De Lara. Orthant-strictly monotonic norms, graded sequences and generalized top-k and k-support norms for sparse optimization, 2019. preprint.

[2] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm, 2019. preprint.

[3] J.-P. Chancelier and M. De Lara. Variational formulations for the l_0 pseudonorm and applications to sparse optimization, 2019. preprint.

A practical primal-dual interior-point algorithm for nonsymmetric conic optimization

Erling D. Andersen
Mosek APS, Fruebjergvej 3, 2100 Copenhagen O, Denmark

Joint work with *Joachim Dahl*

Keywords : Optimization, convex, conic, interior-point methods

It is well known that primal-dual interior-point algorithms for linear optimization can easily be extended to the case of symmetric conic optimization, as shown by Nesterov and Todd (NT) in their 1997 paper about self-scaled barriers. Although many convex optimization problems can be expressed using symmetric cones then models involving for instance exponential functions do not belong to the class of symmetric conic optimization problems. Therefore, several authors have suggested generalizations of the NT primal-dual interior-point algorithm to handle nonsymmetric cones such as the exponential cone. Based on this work we will present a generalization of the NT algorithm to the case of nonsymmetric conic optimization. Although we have no polynomial complexity proof for the suggested algorithm then it performs extremely well in practice as will be documented with computational results.
To summarize, this presentation should be interesting for users of convex optimization.

A stochastic Levenberg-Marquardt Method Using Random Models

Clément Royer
Université Paris-Dauphine

Joint work with *El Houcine Bergou, Youssef Diouane,* and *Vyacheslav Kungurtsev*

Keywords : Stochastic optimization, Model-based methods, Probabilistic properties.

Globally convergent variants of the Gauss-Newton algorithm are often the preferred methods to tackle nonlinear least squares problems. Among such frameworks, Levenberg-Marquardt and trust-region methods are two well-established, similar paradigms, with analyses that are often close in spirit. Both schemes have been successfully studied when the Gauss-Newton model is replaced by a random model, only accurate with a given probability. Meanwhile, problems where even the objective value is subject to noise have gained interest, driven by the need for efficient methods in fields such as data assimilation.
In this talk, we describe a stochastic Levenberg-Marquardt algorithm that handles noisy objective function values and random models, provided sufficient accuracy is achieved in probability. In particular, if the probability of accurate function estimates and models is sufficiently large, we establish that the proposed algorithm converges globally to a first-order stationary point of the problem with probability one. Furthermore, we bound the expected number of iterations needed to reach an approximate stationary point. Our method relies on both a specific scaling of the regularization parameter and a measure of criticality tailored to least-squares problems. We finally present an application of our method to variational data assimilation, where stochasticity arises from the so-called ensemble methods.

Stability of marketable payoffs with re-trading

Jean-Marc Bonnisseau
Paris School of Economics, Université Paris 1 Panthéon-Sorbonne

Joint work with *Achis Chéry*

Keywords : Mathematical economics, financial equilibrium, long-term assets, re-traded assets, stability

We consider a stochastic financial exchange economy with a finite date-event tree representing time and uncertainty and a financial structure with possibly long-term assets. We address the question of the stability of financial structures, that is the continuity of the set of marketable payoffs with respect to the asset prices. In a previous paper, we have exhibited a sufficient condition, which is based only on the return of the assets. However, it is never satisfied in the structures with re-trading (See Bonnisseau-Chéry), which is a very common feature in many papers following the model presented in the book of Magill and Quinzii. The main purpose of this paper is to deepen the study of the stability of financial structures with long-term assets and especially to address this issue in cases where there is a re-trading of assets after their issuance at different nodes and different dates. We exhibit a new sufficient condition on general financial structures, which enjoys the property to be inherited by the re-trading extension of a financial structure. We also show that the situation is more complex when all initial assets are not issued at the initial date and we then provide a stronger condition to cover this case.

References

[1] M. Magill and M. Quinzii, *Theory of Incomplete Markets*, Cambridge University Press (1996)

[2] J-M. Bonnisseau and A. Chéry, *Stability of marketable payoffs with long-term assets*, Annals of Finance, 10 (2014), pp. 523-552.

Tropical dynamic programming for stochastic optimal control

Benoît Tran
École des Ponts ParisTech - CERMICS

Joint work with *Marianne Akian* and *Jean-Philippe Chancelier*

Keywords : Dynamic Programming, Stochastic Optimal Control, Tropical Algebra

We present an algorithm which builds upper or lower approximations of the (Bellman) value functions of a risk-neutral stochastic optimal control problem in discrete time, with white noises of finite supports.

Upper (resp. Lower) approximations of a given value function are built as min-plus linear (resp. max-plus linear) combinations of "basic functions". At each iteration, we add a new, randomly selected, basic functions to the current combination. We give sufficient conditions to ensure asymptotic almost sure convergence of the generated approximating functions to the value functions on sets of interest.

We illustrate this algorithm numerically on optimal control problems with linear dynamics and polyhedral costs.

References

[1] M. Akian, J-P. Chancelier and B. Tran, *A stochastic algorithm for deterministic multistage optimization problems*, preprint, 2019, 35 pages.

[2] M. Akian, J-P. Chancelier and B. Tran, *A Min-plus-SDDP Algorithm for Deterministic Multistage Convex Programming*, Proceedings of the 58th IEEE Conference on Descision and Control, 2019, 6 pages.

Optimization of Palliative Cancer Therapy: the Logic of Containing Tumors

Yannick Viossat
CEREMADE, Université Paris-Dauphine, PSL

Joint work with *Robert Noble*

Keywords : Mathematical oncology, Optimal control, Pharmacoresistance, Adaptive therapy

Many cancer treatments are administered at the maximal tolerated dose, and eventually fail as the tumor becomes resistant. When a cure is highly unlikely, an alternative is to use the minimal effective dose, that is, the minimal dose that allows to contain the tumor while ensuring a sufficient quality of life to the patient. The hope is to diminish both treatment toxicity and selection for resistance to treatment, allowing us to control the tumor for a longer time. Experimental results are promising.
We study how to maximize the time at which tumor size becomes higher than a given threshold. For simple models with two types of cells, sensitive and fully resistant to treatment, we find that under very general conditions, containment strategies are optimal, or almost optimal. This unifies a number of previous approaches. The clinical gains with respect to more aggressive treatments crucially depends on the tumor growth model. In particular, our results are much more robust to modifications of the basic model for Gompertzian growth than for logistic growth. Finally, if resistant cells retain some sensitivity to treatment, the optimal strategy is to first contain the tumor with a low-dose treatment but eventually switch to maximal tolerated dose. This is the opposite of what other authors have found, and we will explain why.
This is part of the PGMO project “Optimization of a new type of cancer therapy”.

EMSx: an Energy Management System numerical benchmark

Adrien Le Franc
CERMICS ENPC

Joint work with *Pierre Carpentier, Jean-Philippe Chancelier,* and *Michel De Lara*

Keywords : Multistage stochastic optimization, Electric microgrid, Numerical benchmark

Inserting renewable energy in the electric grid in a decentralized manner is a key challenge of the energy transition. However, at local stage, both production and demand display erratic behavior, be it for the dependence upon local weather conditions. Thus, the match between supply and demand is delicate to achieve in such a stochastic context. It is the goal of Energy Management Systems (EMS) to achieve such balance at least cost. We present EMSx, a numerical benchmark for testing control algorithms for the management of electric microgrids doted with a photovoltaic unit and an energy storage system. Our benchmark is based on a rich collection of historical observations and forecasts collected by the company Schneider Electric. Besides the dataset, we also publicly release `EMSx.jl`, a package implemented in the Julia language which enables to easily implement a microgrid controller and to test it on the EMSx benchmark. Eventually, we showcase the results of standard microgrid control methods, including Model Predictive Control, Open Loop Feedback Control and Stochastic Dynamic Programming.

Convergence and Dynamical Behavior of the ADAM Algorithm for Non Convex Stochastic Optimization

Anas Barakat
Télécom Paris, Institut Polytechnique de Paris

Joint work with *Pascal Bianchi*

Keywords : Stochastic approximation with constant step, Dynamical systems, Weak convergence of stochastic processes, Kurdyka-Łojasiewicz inequality

Adam is a popular variant of the stochastic gradient descent for finding a local minimizer of a function. The objective function is unknown but a random estimate of the current gradient vector is observed at each round of the algorithm.

1 Dynamical Behavior and Asymptotic Analysis

Assuming that the objective function is differentiable and non-convex, we establish the convergence in the long run of the iterates to a stationary point. The key ingredient is the introduction of a continuous-time version of Adam, under the form of a non-autonomous ordinary differential equation. The existence and the uniqueness of the solution are established, as well as the convergence of the solution towards the stationary points of the objective function. The continuous-time system is a relevant approximation of the Adam iterates, in the sense that the interpolated Adam process converges weakly to the solution to the ODE.

2 Convergence Rates of a Clipped Version of Adam

In this second part, we study the Adam algorithm for smooth nonconvex optimization under a boundedness assumption on the adaptive learning rate. The bound on the adaptive step size depends on the Lipschitz constant of the gradient of the objective function and provides safe theoretical adaptive step sizes. Under this boundedness assumption, we show a novel first order convergence rate result in both deterministic and stochastic contexts. Furthermore, we establish convergence rates of the function value sequence using the Kurdyka-Łojasiewicz property.

HJB equations on stratified domains

Othmane JERHAOUI
ENSTA PARIS/UMA

Joint work with *Hasnaa ZIDANI*

Keywords : HJB Equations, Optimal Control, Nonsmooth Analysis.

In this talk, we are interested in some control problems on stratified domains. In such problems, the state variable space is partitioned in different open sets separated by an interface, that is a collection of lower dimensional manifolds. Each open region is associated with a compact control set, a controlled dynamics and a cost function. The simplest example of such structures (in dimension 2) is two open and disjoint half planes with a common boundary which is a line, see [1, 3].
We are interested in characterizing the value function of the optimal control problem as solution to an adequate Hamilton-Jacobi equation. More precisely, we prove that the essential Hamiltonian used in [4] and [2] characterizes both the sub and super solutions. Finally, we show that a *strong* comparison principle holds even when the solution in the viscosity sense (the value function) is only lower semicontinuous.

References

[1] Guy Barles, Ariela Briani, and Emmanuel Chasseigne. A bellman approach for regional optimal control problems in r^n. *SIAM Journal on Control and Optimization*, 52(3):1712–1744, 2014.

[2] Daria Ghilli, Zhiping Rao, and Hasnaa Zidani. Junction conditions for finite horizon optimal control problems on multi-domains with continuous and discontinuous solutions. *ESAIM: Control, Optimisation, and Calculus of Variations*, 2019.

[3] Zhiping Rao, Antonio Siconolfi, and Hasnaa Zidani. Transmission conditions on interfaces for hamilton–jacobi–bellman equations. *Journal of Differential Equations*, 257(11):3978–4014, 2014.

[4] Zhiping Rao and Hasnaa Zidani. Hamilton–jacobi–bellman equations on multi-domains. In *Control and Optimization with PDE Constraints*, pages 93–116. Springer, 2013.

Optimistic planning algorithm for some state-constrained optimal control problems

Nidhal GAMMOUDI
UMA-ENSTA Paris

Joint work with *Olivier Bokanowski* and *Hasnaa Zidani*

Keywords : Optimal Control, State constraints, Optimistic Optimization, Planning

In this work, we aim at extending reinforcement learning (RL) methods to some state-constrained nonlinear optimal control problems. An important feature of these methods is that they do not need to discretize the state space and their complexity do not grow with the dimension of the state vector. In [2, 3], the RL methods have been analysed for control problems without state constraints.

In our work, we extend the RL to control problems in presence of ponctual state constraints. For these problems, it is known that the value function may be discontinuous. However, following [1], one can describe precisely the value function and its epigraph by using an auxiliary control problem whose value function is Lipschitz continuous. We first extend the optimistic planning algorithm to solve the auxiliary control problem. Then we use this approach to get an approximation of the original control problem with state-constraints and without any controllability assumption. Furthermore, we establish some convergence guarantees and complexity results for our proposed optimistic planning algorithm similar to those presented in [2, 3].

Finally we illustrate our method on some nonlinear optimal control problems.

References

[1] A. Altarovici, o. Bokanowski and H. Zidani, *A general Hamilton-Jacobi framework for non-linear state-constrained control problems*, ESAIM: Control, Optimisation and Calculus of Variations, 2013, pp. 337–357.

[2] L. Buşoniu, E. Páll and R. Munos, *Discounted near-optimal control of general continuous-action nonlinear systems using optimistic planning*, 2016 American Control Conference (ACC), 2016, pp. 203–208.

[3] L. Buşoniu, E. Páll and R. Munos, *Continuous-action planning for discounted infinite-horizon nonlinear optimal control with Lipschitz values*, Automatica, 2018, pp. 100–108.

Algorithmic, Combinatorial, and Geometric Aspects of Linear Optimization

Antoine Deza
McMaster University, Ontario, Canada

Keywords : Optimization Algorithms, Combinatorial Optimization, Discrete Geometry

The simplex and interior point methods are currently the most computationally successful algorithms for linear optimization. While the simplex methods follow an edge path, the interior point methods follow the central path. The algorithmic issues are closely related to the combinatorial and geometric structure of the feasible region. Focusing on the analysis of worst-case constructions leading to computationally challenging instances, we discuss connections to the largest diameter of lattice polytopes, to the complexity of convex matroid optimization, and to the number of generalized retarded functions in quantum field theory. Complexity results and open questions are also presented. In particular, we answer a question raised in 1986 by Colbourn, Kocay, and Stinson by showing that deciding whether a given sequence is the degree sequence of a 3-hypergraph is computationally prohibitive.

References

[1] A. Deza, A. Levin, S. Meesum, and S. Onn, *Optimization over degree sequences*, SIAM Journal on Discrete Mathematics (2018), pp. 2067–2079.

[2] A. Deza, G. Manoussakis, and S. Onn, *Primitive zonotopes*, Discrete and Computational Geometry (2018), pp. 27–39

[3] A. Deza, L. Pournin, and N. Sukegawa, *The diameter of lattice zonotopes*, Proceedings of the American Mathematical Society (to appear).

A nonlinearly-constrained bundle method for nonconvex chance-constrained problems

Paul Javal
EDF R&D and CMA, MINES ParisTech, PSL Research University

Joint work with ***W. van Ackooij, S. Demassey, H. Morais, W. de Oliveira*** **and** ***B. Swaminathan***

Keywords : Nonconvex Optimization, Stochastic Programming, Nonsmooth Optimization

We consider a class of nonconvex chance-constrained optimization problems whose involved functions can be represented by the Difference of Convex (DC) functions. A DC function approximating the probability constraint is investigated, and the resulting DC-constrained optimization problem is handled by a new infeasible bundle method based on the so-called improvement functions. The proposed algorithm neither employs penalization techniques nor solves subproblems with linearized constraints. The approach is assessed in a class of several (nonconvex) chance-constrained programs.

Linear Bellman operators with stochastic costs preserves polyhedrality

Maël Forcier
École des Ponts, École polytechnique

Joint work with *Stéphane Gaubert,* and *Vincent Leclère*

Keywords : Stochastic programming, Polyhedra, Dynamic Programming, Normal fan

In the current context of climate change and energy transition, the need of renewable energies is growing rapidly. Dealing with intermittent and hard to predict sources of energies such as wind and solar is a challenge for the electric system. The field of stochastic programming studies algorithms which account for this impredictability by optimizing in an uncertain environment. A particular case is multistage stochastic programming which considers problems with a finite number of time steps. At each time-step some random event occurs and the Decision Maker has to take a decision based on past events. Such problems can be solved by dynamic programming methods. One can define the value function at each time step and then write a Bellman equation to solve the problem with a backward recursion. Solving the multistage problem becomes then a problem of computing and/or approximating these values functions. In particular, methods of approximation by adding cuts (i.e affine lower approximations), such as Benders decomposition or Stochastic Dual Dynamic Programming (SDDP), are well adapted to the case where all the value functions are polyhedral (i.e convex piecewise affine). In the linear setting, if further all the stochastic parameters have a discrete distributions, it is well known that all value functions are polyhedral. In this talk, we prove that if the linear costs in the objective function are the only stochastic parameters, then the value functions are still polyhedral. The proof relies on a geometric representation of optimal actions in terms of a normal fan to a polyhedron. We also give counter examples showing that the result does not hold if the constraint parameters are random.

1 Definitions, notations and recalls

2 Preserving polyhedrality with stochastic cost

2.1 Reduction to a finite sum with normal fan

2.2 Piecewise affine and polyhedral function

3 Counter examples with a stochastic constraint

References

[1] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*, SIAM, (2009).

[2] T. Rockafellar and R. Wets, *Variational analysis*, Springer Science & Business Media, 317 (2009).

[3] S. Lu and S. Robinson, *Normal fans of polyhedral convex sets*, Set-Valued Analysis, 16 (2008), pp. 281–305.

Unconstrained relaxations and duality in global optimization

Simon Boulmier
LocalSolver

Keywords : Global optimization, nonlinear programming, augmented Lagrangian algorithm, Wolfe duality, unconstrained optimization.

The global optimization of mixed-integer nonlinear programs is an ambitious goal, with many practical applications. Here, we consider the broadest class of such problems, where no convexity or regularity assumptions are made. We only require that an expression graph representing the objective and constraints be given.
The most common approach to solve these generic problems is the *reformulation convexification* technique. It uses a symbolic reformulation to express the problem in a standard form, such as the one introduced in [1], that facilitates the implementation of various tools like convex relaxations and bound tightening techniques. Some of the most successful global solvers [2, 3] follow this approach and make use of decades of engineering in mixed-integer linear programming for robustness and efficiency.
Since the work of [4], we know that the hybridation with nonlinear programming can lead to important improvements. However, linear relaxations still rule in most global optimization solvers and remain the default setting. To show that nonlinear relaxations can be competitive, we implemented a global optimization solver that can use both linear and nonlinear relaxations. This presentation will be focused on the use of unconstrained optimization in this context, a key idea that substantially improved both the performance and robustness of the solver.

References

[1] E. Smith, C. Pantelides, A symbolic reformulation algorithm for the global optimisation of nonconvex MINLPs, *Computers & Chemical Engineering*, vol. 23:457–478, 1999.

[2] N. Sahinidis, M. Tawarmalani, Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications, *Springer Science & Business Media*, vol. 65, 2002.

[3] S. Vigerske, A. Gleixner, SCIP: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework, *Taylor & Francis*, 33(3):563–593, 2018.

[4] A. Khajavirad, N. Sahinidis, A hybrid LP/NLP paradigm for global optimization relaxations, *Mathematical Programming Computation* 10(3):383–421, 2018.

Long term dynamics of the subgradient optimization method in the non-smooth and non-convex case

Rodolfo Ríos-Zertuche
Artificial and Natural Intelligence Toulouse Institute

Keywords : Optimization, subgradient method, closed measures

The advent of huge scale non convex non smooth and poorly structured optimization problems (e.g. deep learning) has triggered a revival of interest in the non-smooth explicit subgradient method of Shor. This problem appears to be much more delicate than implicit versions (like the proximal gradient and PALM algorithms). It features indeed highly oscillatory behavior even with rigidity assumptions like semi-algebraicity of the objective function. In recent work, we analyze the algorithm under fairly weak assumptions and we prove some convergence properties. To do this we use closed measures, a tool taken from geometric measure theory that we expect to be useful also for the analysis of convergence of other algorithms involving sequences in Euclidean spaces.

Multi-objective infinite horizon optimal control problems: characterization of the Pareto fronts and reconstruction of optimal trajectories

Ana Paula Chorobura
ENSTA Paris - France

Joint work with *Hasnaa Zidani*

Keywords : Multi-objective programming; Infinite horizon optimal control problem; Trajectory reconstruction; Semi-Lagrangian scheme.

We will present a method to compute the Pareto fronts of multi-objective infinite horizon optimal control problems with state constraints. The approach is based on the idea introduced in [2] for the finite horizon case. First, a mono-objective auxiliary optimal control problem, free of state constraints, is introduced. The zero level set of the value function of the auxiliary control problem is related with the (weak) Pareto front of the multi-objective optimal control problem. Moreover, a more detailed characterization of the Pareto front for the multi-objective infinite horizon control problem is given. In the infinite horizon context, the value function of the auxiliary optimal control problem satisfies a Hamilton-Jacobi-Bellman equation, however it is not the unique solution. Usually numerical approximations are analyzed (convergence and error estimate) under the assumption that a comparison principle holds. Based on [3] a Semi-Lagrangian method is proposed to compute the value function of the auxiliary control problem and the proof of convergence results is obtained by using only the Dynamic Programming Principle. Furthermore, we discuss an extension of the trajectory reconstruction algorithm [1] to the case of Pareto solution of infinite horizon control problems

References

[1] M. Assellaou, O. Bokanowski, A. Désilles, and H. Zidani. Value function and optimal trajectories for a maximum running cost control problem with state constraints. Application to an abort landing problem. *ESAIM:M2AN*, 52(1):305–335, 2018.

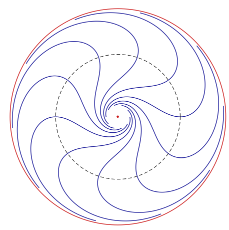
[2] A. Désilles and H. Zidani. Pareto front characterization for multi-objective optimal control problems using Hamilton-Jacobi approach. *SIAM J. on Control and Optimization*, 57(6):3884–3910, 2019.

[3] M. Falcone and R. Ferretti. *Semi-Lagrangian approximation schemes for linear and Hamilton-Jacobi equations.* SIAM, Philadelfia, 2014.

Abnormal Trajectory in Optimal Control with Application in Hydrodynamics and Geometry

Boris Wembe
IRIT-UPS, Toulouse, France

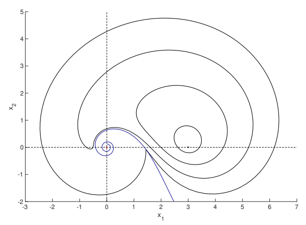
Keywords: N-Vortex problem, Finslerian Geometry, Reeb foliation, optimal synthesis, cut locus.



Reeb's circle

We consider the 2-vortex problem in the plane whose dynamics is $\dot{q}(t) = F_0(q(t)) + u_1(t)\,F_1(q(t)) + u_2(t)\,F_2(q(t))$, where $q := (x, y) \in M := \mathbb{R}^2 \setminus \{0\}$ is a passive particle moving under the influence of a vortex, where the current $F_0 := \frac{\mu}{x^2+y^2}(-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y})$ can be, at q, weak ($\|F_0(q)\| < 1$), strong ($\|F_0(q)\| > 1$) or intermediate ($\|F_0(q)\| = 1$), where $F_1 := \frac{\partial}{\partial x}$, $F_2 := \frac{\partial}{\partial y}$ define the control directions and where $u := (u_1, u_2) \in \mathcal{U} = L^\infty(\mathbb{R}_+, B(0,1))$ is the control. For $q_0 \in M$ given, we propose to construct the time-minimal synthesis, that is to study the value function: $T(q_0, q_1) := \inf_{u \in \mathcal{U}} t_f$ s.t $q(t_f, q_0, u) = q_1$.

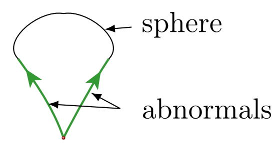
First, we study the extremal flow. Thanks to the integrability of the system and a rotational symmetry, we construct a Reeb's component, that emphasizes the diaphragm shape of the flow, in order to classify the extremals.



Level sets: weak case

Then, for the synthesis, we distinguish the weak and strong current cases. In the weak case we are reduced to a Randers metric, the Finslerian framework then allows us to characterize the cut locus and to construct the synthesis. In this case, the minimum time function is continuous and there are no abnormal extremals. On the other hand, the strong case goes beyond the Finslerian framework because there are minimizing abnormals that induce the loss of continuity of the value function and give a fan shape to the short time balls. In this work, we establish two results, a first one on the continuity of the value function and a second on the characterization of the cut locus which are essential to construct the optimal synthesis.

Besides, this toy model classifies the role of abnormal trajectories in optimal control and geometry. Furthermore, it is the starting point to analyze the general cases with several vortices which, in the Hamiltonian frame [4], has applications in hydrodynamics and in the N-body problem.



Small ball: strong case

References

[1] B. Bonnard, O. Cots, & B. Wembe, *A Zermelo navigation problem with a vortex singularity.* Submitted September 2019.

[2] B. Bonnard & I. Kupka, *Théorie des singularités de l'application entrée/sortie et optimalité des trajectoires singulières dans le problème du temps minimal*, Forum Math., **5** (1993), no. 2, pp. 111–159.

[3] D. Bao, S.-S. Chern & Z. Shen, *An Introduction to Riemann-Finsler Geometry*, vol 200 of Graduate Texts in Mathematics, Springer-Verlag New York, 2000, 435 pages.

[4] K. Meyer, G. Hall & D. C. Offin, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, vol. 90 of *Applied Mathematical Sciences*, Springer-Verlag New York, 2009, 399 pages.

Solving Perfect Information Mean Payoff Zero-sum Stochastic Games by Variance Reduced Deflated Value Iteration

Omar Saadi
CMAP, École polytechnique and INRIA

Joint work with *Marianne Akian, Stéphane Gaubert,* and *Zheng Qu*

Keywords : Mean Payoff Stochastic Games, Deflated Value Iteration, First Hitting Times, Variance Reduction.

We introduce a deflated version of value iteration, which allows one to solve mean-payoff problems, including both Markov decision processes and perfect information zero-sum stochastic games. This method requires the existence of a distinguished state which is accessible from all initial states and for all policies; it differs from the classical relative value iteration algorithm for mean payoff problems in that it does not need any primitivity or geometric ergodicity condition. Our method is based on a reduction from the mean payoff problem to a discounted problem by a Doob h-transform, combined with a deflation technique and non-linear spectral theory results (Collatz-Wielandt characterization of the eigenvalue), inspired by [1]. In this way, we provide a new method *Deflated Value Iteration* that allows to extend complexity results from the discounted to the mean payoff case. In particular, Sidford, Wang, Wu and Ye [2] developed recently an algorithm combining value iteration with variance reduction techniques to solve discounted Markov decision processes in sublinear time when the discount factor is fixed. We combine deflated value iteration with variance reduction techniques to obtain sublinear algorithms for mean payoff stochastic games in which the first hitting times of a distinguished state are bounded a priori.
A first account of the present work can be found in [3].

References

[1] M. Akian and S. Gaubert, *Policy iteration for perfect information stochastic mean payoff games with bounded first return times is strongly polynomial*, *arXiv preprint arXiv:1310.4953*, 2013.

[2] A. Sidford, M. Wang, X. Wu and Y. Ye, *Variance reduced value iteration and faster algorithms for solving Markov decision processes*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 2018, pp. 770–787.

[3] M. Akian, S. Gaubert, Z. Qu and O. Saadi, *Solving Ergodic Markov Decision Processes and Perfect Information Zero-sum Stochastic Games by Variance Reduced Deflated Value Iteration*, *arXiv preprint arXiv:1909.06185*, 2019.

Optimization in Repeated Mechanisms: Challenges

Clément Calauzènes
Criteo AI Lab

Keywords : Machine learning, Web advertising, Auctions Theory

Online Advertising heavily relies on auctions to sell advertising space. In this context, sellers (Ad Exchanges) and buyers (advertisers) repeatedly play one against each other at the rate of billions of auctions a day. In order to learn policies to play this game, not only the players (bidders and seller) need to formalize well the behavior of their opponents, but also face complex optimization problems: adversarial by definition, online by constraint.

Optimal classifiers for adversarial settings under uncertainty

Benjamin Roussillon
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, DGA

Joint work with *Patrick Loiseau*

Keywords : Games, Optimization, Classification, Adversarial learning

1 Model

We consider the problem of finding optimal classifiers in an adversarial setting where a defender observes the behavior of users through a feature vector v with a finite set of possible values $\mathcal{V}$ and must decide whether this behavior is malicious (class-1) or non-malicious (class-0). A vector is considered malicious if it comes from an attacker (which aims to maximize its utility) and non-malicious if it is comes from a normal user (which behavior is described with a probability distribution over feature vectors). Such a setting can be seen for example in credit card fraud where a given transaction can either be malicious if it comes from a thief (who then gains money if the transaction is successful) or non-malicious if it comes from a client of the bank. Unlike other models in the literature we deal with an attacker whose objective is not known to the defender.
To model this situation, we propose a Bayesian game framework inspired from [1] where the defender chooses a classifier with no *a priori* restriction on the set of possible classifiers. The key difficulty in the proposed framework is that the set of possible classifiers is exponential in the number of possible attacks.

2 Results

We first prove that the strategy set of the defender can be effectively reduced by computing the optimal *probability of detection* for each vector instead of the full strategy of the defender. Such reductions are commonly found in security games and would allow us to find the Bayesian Nash equilibrium in polynomial time. However, the number of possible attacks can itself be exponentially large (relative to the number of features used for classification). To counter this, we show that the equilibrium can be characterized completely via functional threshold classifiers with a small number of parameters (depending only on the number of possible types of attackers) which:

- Can be computed in polynomial time in the number of attacks.
- Can be approximated with limited knowledge of the parameters of the game.
- Can be trained on subsets of $\mathcal{V}$ with provable efficiency.
- Can be used in models where $\mathcal{V}$ is continuous.

References

[1] Lemonia Dritsoula, Patrick Loiseau, and John Musacchio. A game-theoretic analysis of adversarial classification. *IEEE Transactions on Information Forensics and Security*, 12(12):3094–3109, December 2017.

Calcul de bornes dans LocalSolver 9.5

Nikolas Stott
LocalSolver, 36 avenue Hoche, Paris, France

Keywords : Solveur, Optimisation globale, Variables mixtes

1 Contexte

LocalSolver est un solveur d'optimisation mathématique tout terrain qui permet de résoudre des problèmes multi-objectifs modélisés avec les variables classiques de la programmation linéaire en nombres entiers (booléens, entiers et réels), ainsi que des variables de type collection (listes et ensembles) pour modéliser par exemple des problèmes de tournées de véhicules et de rangement. La recherche de solution repose sur le coeur historique de méthodes heuristiques, en utilisant des techniques de recherche locale, et sur un module d'optimisation globale pour traiter les problèmes de type MINLP.
L'objectif de LocalSolver est de produire des solutions de qualité rapidement sur tout l'éventail des problèmes modélisables. Ces solutions sont ensuite raffinées. Afin d'attester de la qualité des solutions obtenues, LocalSolver calcule des bornes inférieures (resp. supérieures) pour chacun des objectifs minimisés (resp. maximisés), et la même stratégie est adoptée: obtenir de premières bornes de qualité rapidement, puis les améliorer itérativement.

2 Techniques de calcul de bornes

LocalSolver représente un problème d'optimisation sous la forme d'un graphe dirigé acyclique. Les bornes associées à chaque noeud du graphe peuvent être déduites des bornes de ses enfants, ou inférées à partir des bornes de ses parents. Cette arithmétique d'intervalle permet d'obtenir des premières bornes sur tous les objectifs dès le pré-processing du modèle.
Dans le cas d'un seul objectif, le module d'optimisation globale reformule le problème et produit des bornes valides au cours de la résolution à l'aide de relaxations linéaires/convexes couplées avec un mécanisme de *branch-and-reduce*. Si le problème est purement linéaire, il est résolu directement par le solveur linéaire sous-jacent.
Dans le cas multi-objectif, le calcul des bornes traite les objectifs un par un. Chaque objectif doit être résolu jusqu'à l'optimum global en contraignant chacun des objectifs précédents à leur valeur optimale avant de traiter le suivant.
Le sous-solveur MINLP permet également d'obtenir des solutions et des bornes lorsque les variables de décisions sont des ensembles, par linéarisation booléenne exacte de la structure ensemblistes et des contraintes associées.
Les problèmes de tournées sont modélisés avec des listes dans LocalSolver. Le calcul de bornes utilisant seulement des arguments de propagation obtient des bornes trop faibles, et il est donc traité avec une technique dédiée: la résolution de la relaxation lagrangienne de Held-Karp pour les TSP symétriques et asymétriques.

3 Résultats

LocalSolver trouve une solution optimale à 40% des problèmes de la MINLPlib et des solutions faisables à 70% d'entre eux. Nous illustrons également un problème industriel, sur lequel la résolution à l'optimum des premiers objectifs entraîne une amélioration significative sur la résolution des objectifs suivants.

The evolutionary dynamics of costly signaling

Christina Pawlowitsch
Université Panthéon-Assas, Paris II

Joint work with *Josef Hofbauer*

Keywords : Games

Costly-signaling games have a remarkably wide range of applications, from education as a costly signal in the job market over handicaps as a signal for fitness in mate selection to politeness in language. While the use of game-theoretic equilibrium analysis in such models is often justified by some intuitive dynamic argument, the formal analysis of evolutionary dynamics in costly-signaling games has only recently gained more attention. In this paper, we study evolutionary dynamics in two basic classes of games with two states of nature, two signals, and two possible reactions in response to signals: a discrete version of Spence's (1973) model and a discrete version of Grafen's (1990) formalization of the handicap principle. We first use *index theory* to give a rough account of the dynamic stability properties of the equilibria in these games. Then, we study in more detail the replicator dynamics and to some extent the best-response dynamics. We relate our findings to equilibrium analysis based on classical, rationality-oriented methods of equilibrium refinement in signaling games.

References

[1] A. Grafen, *Biological signals as handicaps*, Journal of Theoretical Biology, 144 (4) (1990), pp. 517–546.

[2] M. Spence, *Job market signaling*, Quarterly Journal of Economics , 87 (3): pp. 355–374.

Sustainable Equilibria

Rida Laraki
CNRS, Lamsade (Dauphine-PSL) and University of Liverpool

Joint work with *Srihari Govindan* and *Lucas Pahl*

Keywords : Nash Equilibrium, Index Theory, Hopf Extension Theorem, Delaunay Triangulation

Following the ideas laid out in Myerson [2], Hofbauer [1] defined an equilibrium of a game as sustainable if it can be made the unique equilibrium of a game obtained by deleting a subset of the strategies that are inferior replies to it, and then adding others. Hofbauer also formalized Myerson's conjecture about the relationship between the sustainability of an equilibrium and its index: for generic games, an equilibrium is sustainable iff its index is +1. von Schemde and von Stengel [3] proved this conjecture for bimatrix games. This paper shows that the conjecture is true for all finite games.

References

[1] Hofbauer, J. (2000): "Some Thoughts on Sustainable/Learnable Equilibria," Mimeo.

[2] Myerson, R.B. (1996): "Sustainable Equilibria in Culturally Familiar Games," in *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, edited by W. Albers et al, Springer, 111–21.

[3] von Schemde, A., and B. von Stengel (2008): "Strategic Characterization of the Index of an Equilibrium," SAGT 2008, LNCS 4997, 242–54.

A stochastic Lanvegin sampling approach to Median of Means estimation

Stéphane Chrétien
Université Lyon 2

Keywords : Median of Means estimation, Langevin dynamics

Big Data are at the heart of all substantial progress in current machine learning with impact on self driving cars, softwares outperforming human efficiency standards, decision making in medicine, law, chess or Go. Nevertheless, data from real-world experiments oftentimes tend to be corrupted with outliers and/or exhibit heavy tails. For instance, data from finance, often suffer from heavy-tailed phenomena. Also, gene expression data are usually not clean and may be corrupted by the presence of outliers. Fraud detection is another form of outlier detection.
Catoni's program on robustness. In order to address these challenging data analytics problems, there is an urgent need for methods that automatically correct for such undesirable phenomena. This program was initiated by Catoni for the estimation of the mean on the real line and Audibert and Catoni in the least-squares regression framework. To the best of our knowledge, estimation of the mean of a real valued random variable is still best performed using the Median-of-Means (MoM) approach. In regression models, Median of Means can handle outliers or heavy tails in the covariates as well as in the reponse.
The MoM estimator. One usually defines the Median of Means estimator of associated with any function f, any permutation π and any number of groups $K|N$, as

$$\mathrm{MoM}_K^\pi[f] = \text{median} \left(\left\{ \frac{1}{N/K} \sum_{i=(k-1)N/K+1}^{kN/K} f(X_{\pi(i)}) \right\}_{k=1,\dots,K} \right). \tag{1}$$

Let $X_1,\dots,X_N$ denote a sample from distribution P, taking values in $\mathbb{R}^d$. Statistical estimation is usually performed via solving

$$\min_{f\in\mathcal{F}} \ \mathrm{MoM}_K^\pi[f]. \tag{2}$$

Since early work of Catoni, Lerasle, Oliveira, Lugosi and Mendelson among others, MoM estimators have been extensively studied lately. See e.g. [2] and the references therein. The only existing result on the computability of MoM estimators is proposed in [1]. However this last work's proposal is based on SDP relaxation of Sum of Squares certificates and the resulting approach is not scalable using nowaday's available toolboxes.

Our results. In the present work, we propose a new scalable approach to MoM estimation based on resampling at each iteration of a stochastic Langevin sampler. We prove polynomial time convergence with high probability based on hitting time analysis and give new underpinnings to some observations in [2].

References

[1] Hopkins, S.B., *Sub-Gaussian mean estimation in polynomial time.* arXiv preprint arXiv:1809.07425.

[2] Lecué, G. and Lerasle, M., *Robust machine learning by median-of-means: theory and practice.* arXiv preprint arXiv:1711.10306.

Paiement d'équilibres algébriques de jeux finis à paiement entier.

Guillaume Vigeral
CEREMADE, Université Paris-Dauphine

Keywords : Jeux, équilibres, paiements d'équilibre

Tout jeu (statique) fini à paiement entier a, par application du théorème de Tarski Seidenberg, au moins un paiement d'équilibre dont toutes les composantes sont des nombres algébriques. On montre la réciproque suivante : si N est supérieur ou égal à 3 et e est un N-uplet de nombres algébrique, il existe un jeu fini à paiement entier et à N joueurs ayant un unique équilibre de Nash, de paiement e.
Une conséquence est que, pour N supérieur ou égal à 3, tout ensemble compact non vide de R^N défini par un nombre fini d'union et intersection d'inégalités polynomiales à coefficients entiers est l'ensemble des paiements d'équilibre d'un jeu à N joueurs et à paiement entier.
On montre également des résultats similaires concernant les équilibres plutôt que les paiements d'équilibres.
Tout jeu (statique) fini à paiement entier a, par application du théorème de Tarski Seidenberg, au moins un paiement d'équilibre dont toutes les composantes sont des nombres algébriques. On montre la réciproque suivante : si N est supérieur ou égal à 3 et e est un N-uplet de nombres algébrique, il existe un jeu fini à paiement entier et à N joueurs ayant un unique équilibre de Nash, de paiement e.
Une conséquence est que, pour N supérieur ou égal à 3, tout ensemble compact non vide de R^N défini par un nombre fini d'union et intersection d'inégalités polynomiales à coefficients entiers est l'ensemble des paiements d'équilibre d'un jeu à N joueurs et à paiement entier.
On montre également des résultats similaires concernant les équilibres plutôt que les paiements d'équilibres.

Long Information Design

Tristan Tomala
HEC Paris

Joint work with *Frédéric Koessler, Marie Laclau* and *Jérôme Renault*

Keywords : Dynamic Games, Information Design, Concavification, functional equations

We analyze zero-sum games between two designers who provide information to a decision maker (DM). Before the DM takes its decision, designers repeatedly disclose information about fixed state parameters. Designers face admissibility constraints on distributions of posterior beliefs they can induce. Our main results characterize equilibrium payoffs and strategies for various timings of the game: simultaneous or alternating disclosures, with or without deadline.

We define concave and convex closure of functions of beliefs along correspondences of admissible distributions of posterior. We show that the value of the game without deadline is the unique solution v of a Mertens-Zamir system of equations

$$v(p,q) = \underset{p}{\mathrm{cav}} \min(u,v)(p,q) = \underset{q}{\mathrm{vex}} \max(u,v)(p,q)$$

where $u(p,q)$ is the realized payoff as function of the beliefs (p,q) of the DM about the state parameters, cav_p is the concave closure operator with respect to p along the admissibility correspondence, vex_q is the convex closure operator with respect to q along the admissibility correspondence.

We also show that in equilibrium, information is disclosed in a single stage when all distributions of posteriors are admissible. With constraints, there may be no bound on the number of stages used to disclose information in equilibrium.

References

[1] F. Koessler, M. Laclau, J. Renault and T. Tomala, *Long Information Design*, working paper (2019).

[2] R. Laraki and J. Renault, *Acyclic Gambling Games*, Mathematics of Operation Research (2019), forthcoming.

Learning in Dynamic Routing Games with Symmetric Incomplete Information

Emilien Macault
HEC Paris

Joint work with *Marco Scarsini (LUISS)* and *Tristan Tomala (HEC Paris)*

Keywords : Routing games, Repeated games, Incomplete information, Rational learning

We introduce a model of dynamic routing games under symmetric incomplete information. It consists in a routing game in which cost functions are determined by an unknown state of the world, as in [1] and [2], to the difference that the game is played repeatedly in discrete time. At each stage, a new mass of players routes over the network, publicly revealing its realized equilibrium costs. Our objective is to study how information aggregates according to the equilibrium dynamics and to which extent a centralized information system can learn about the state of the world. We define several forms of learning – whether agents eventually learn the state of the world or act as in full information – and provide simple examples showing that in the general case, with a non-atomic set of players, learning may fail and routing may be inefficient even in the looser sense. This contrasts with the atomic case, in which a folk theorem ensures players can learn the game parameters. In a non-atomic setup, learning cannot be ensured unless there is an additional source of randomness to incentivize exploration of the network. We show that this role can be fulfilled by an exogenous source of randomness in the game. We first explore the case of a variable and unbounded demand size. Our main result proves that under some network topology condition as in [3] and costs unboundedness, a variable and unbounded demand is sufficient to ensure learning. This theorem can be adapted whether costs are observed edge by edge or path by path. We further extend this result to infinite state spaces. We additionally provide examples to show these conditions are tight. In a second variant of our model, we connect our work with the behavioural learning and cascade literature *à la* Smith and Sorensen [4] and show that if instead of having random demand size, costs are observed with some unbounded noise, then learning may occur under some conditions on the distribution of noises.

References

[1] D. Acemoglu, A. Makhdoumi, A. Malekian and A. Ozdaglar, *Informational Braess' Paradox: The Effect of Information on Traffic Congestion*, Oper. Res., 66 (2018), pp. 893–917.

[2] W. Manxi, A. Saurabh and A. Ozdaglar, *Value of Information Systems in Routing Games*, ArXiv preprint (2018).

[3] I. Milchtaich, *Network Topology and the Efficiency of Equilibrium*, Games and Economic Behavior, 57 (2006), pp. 321–346.

[4] L. Smith and P. Sørensen, *Pathological Outcomes of Observational Learning*, Econometrica, 68 (2000), pp. 371–398.

This work has been partly supported by COST Action CA16228 European Network for Game Theory.

A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix

Bento Natura
London School of Economics

Joint work with *Daniel Dadush, Sophie Huiberts,* and *László A. Végh*

Keywords : Linear Programming, Interior Point Methods, Layered Least Squares, Linear Matroid

Following the breakthrough work of Tardos (Oper. Res. '86) in the bit-complexity model, Vavasis and Ye (Math. Prog. '96) gave the first exact algorithm for linear programming in the real model of computation with running time depending only on the constraint matrix. For solving a linear program (LP) $\max c^\top x,\ Ax = b,\ x \geq 0,\ A \in \mathbb{R}^{m\times n}$, Vavasis and Ye developed a primal-dual interior point method using a *'layered least squares'* (LLS) step, and showed that $O(n^{3.5}\log(\bar{\chi}_A + n))$ iterations suffice to solve (LP) exactly, where $\bar{\chi}_A$ is a condition measure controlling the size of solutions to linear systems related to A.
Monteiro and Tsuchiya (SIAM J. Optim. '03), noting that the central path is invariant under rescalings of the columns of A and c, asked whether there exists an LP algorithm depending instead on the measure $\bar{\chi}^*_A$, defined as the minimum $\bar{\chi}_{AD}$ value achievable by a column rescaling AD of A, and gave strong evidence that this should be the case. We resolve this open question affirmatively.
Our first main contribution is an $O(m^2n^2+n^3)$ time algorithm which works on the linear matroid of A to compute a nearly optimal diagonal rescaling D satisfying $\bar{\chi}_{AD} \leq n(\bar{\chi}^*)^3$. This algorithm also allows us to approximate the value of $\bar{\chi}_A$ up to a factor $n(\bar{\chi}^*)^2$. This result is in (surprising) contrast to that of Tunçel (Math. Prog. '99), who showed NP-hardness for approximating $\bar{\chi}_A$ to within $2^{\mathrm{poly}(\mathrm{rank}(A))}$. The key insight for our algorithm is to work with ratios g_i/g_j of circuits of A—i.e., minimal linear dependencies $Ag = 0$—which allow us to approximate the value of $\bar{\chi}^*_A$ by a maximum geometric mean cycle computation in what we call the *'circuit ratio digraph'* of A.
In our second main contribution we develop a *scaling invariant* LLS algorithm, which uses and dynamically maintains improving estimates of the circuit ratio digraph, together with a refined potential function based analysis for LLS algorithms in general. With this analysis, we derive an improved $O(n^{2.5}\log n\log(\bar{\chi}^*_A + n))$ iteration bound for optimally solving (LP) using our algorithm. The same argument also yields a factor $n/\log n$ improvement on the iteration complexity bound of the original Vavasis-Ye algorithm.

References

[1] Stephen A. Vavasis and Yinyu Ye, *A primal-dual interior point method whose running time depends only on the constraint matrix*, Mathematical Programming, July 1996, Volume 74, pp. 79–120.

[2] Éva Tardos, *A strongly polynomial algorithm to solve combinatorial linear programs*, Operations Research, 1986, pp. 250 –256.

[3] Levent Tunçel, *Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard"*, Mathematical Programming, 1999, pp. 219 –223.

Mean Field Games with Branching

Julien Claisse
Université Paris–Dauphine

Joint work with *Zhenjie Ren* and *Xiaolu Tan*

Keywords : Mean field games, Branching diffusion process, Relaxed control

Mean field games are concerned with the limit of large-population stochastic differential games where the agents interact through their empirical distribution. In the classical setting, the number of players is large but fixed throughout the game. However, in various applications, such as population dynamics or economic growth, the number of players can vary across time which may lead to different Nash equilibria. For this reason, we introduce a branching mechanism in the population of agents and obtain a variation on the mean field game problem. As a first step, we study a simple model using a PDE approach to illustrate the main differences with the classical setting. Then we study the problem in a general setting by a probabilistic approach, based upon the relaxed formulation of stochastic control problems.

References

[1] J. Claisse, Z. Ren and X. Tan, *Mean Field Games with Branching*, arXiv:1912.11893 (2019), pp. 1–32.

Interior point methods for logistic regression

François Pacaud
Artelys, France, francois.pacaud@artelys.com

Joint work with *Rubing Shen*

Keywords : Non-linear programming, logistic regression, hyperparameters optimization

Logistic regression is a well known statistical model used to predict binary outcomes. Fitting logistic regression models reduces to the minimization of an empirical loss penalizing the gap between the values predicted by the statistical model and the observations. The corresponding optimization problem formulates naturally as a (unconstrained) non-linear problem. Then, the numerical resolution of the problem often fallbacks to a first order method. However, if the size of the dataset is not too large, second order methods also have proven to be effective [3]. In this talk, we study the resolution of logistic regression problems with a commercial interior point solver, Artelys Knitro [2]. We first provide numerical studies showing that Knitro behaves well to fit logistic regression models, both with ℓ_1 and ℓ_2 penalties. We compare Knitro with the solver L-BFGS-B [5], currently used in the library Scikit-Learn [4]. In the second part of the talk, we focus on the optimization of regularization hyperparameters. We suppose given a cross-validation evaluation function, based onto a validation dataset. By using the implicit function theorem, we derive the sensivity of the cross-validation loss function w.r.t. the penalty term [1], thus allowing to optimize it with a descent algorithm based on BFGS. The devised algorithm is faster than classical methods such as grid searches, and outperforms the finite-difference algorithm implemented inside Knitro. Eventually, we depict some ideas to improve interior point algorithms for the resolution of optimization problems arising in machine learning.

References

[1] Yoshua Bengio. Gradient-based optimization of hyperparameters. *Neural computation*, 12(8):1889–1900, 2000.

[2] Richard H Byrd, Jorge Nocedal, and Richard A Waltz. Knitro: An integrated package for nonlinear optimization. In *Large-scale nonlinear optimization*, pages 35–59. Springer, 2006.

[3] Chih-Jen Lin, Ruby C Weng, and S Sathiya Keerthi. Trust region newton method for logistic regression. *Journal of Machine Learning Research*, 9(Apr):627–650, 2008.

[4] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[5] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.

Solving Alternative Current Optimal Power Flow to Global Optimality with Quadratic Reformulation Using Semi-Definite Programming and Branch-and-Bound

Manuel Ruiz
RTE Réseau de Transport d'Electricité, France

Joint work with *Hadrien Godard, Sourour Elloumi, Amélie Lambert* and *Jean Maeght*

Keywords : Optimal Power Flow, Semi-Definite Programming, Branch-and-Bound, Quadratic Reformulation, Global Optimization

Alternative Current Optimal Power Flow (ACOPF) is naturally formulated as a non-convex problem. In that context, solving (ACOPF) to global optimality remains a challenge when classic convex relaxations are not exact. We use semidefinite programming to build a quadratic convex relaxation of (ACOPF). We show that this quadratic convex relaxation has the same optimal value as the classical semidefinite relaxation of (ACOPF) which is known to be tight. In that context, we build a spatial branch-and-bound algorithm to solve (ACOPF) to global optimality that is based on a quadratic convex programming bound. An advantage of our approach is that we capture the strength of semidefinite programming bounds into a quadratic convex program that is known to be faster to solve.

References

[1] W. A. Bukhsh, A. Grothey, K. I. M. McKinnon and P. A. Trodden, Local Solutions of the Optimal Power Flow Problem. *IEEE Transactions on Power Systems*, 28(4):4780–4788, 2013.

[2] A. Billionnet, S. Elloumi, and A. Lambert, Exact quadratic convex reformulations of mixed-integer quadratically constrained problems. *Mathematical Programming*, 158(1):235–266, 2016.

[3] C. Josz, Application of polynomial optimization to electricity transmission networks. *PhD thesis, arXiv:1608.03871*, 2016.

[4] J. Lavaei and S. H. Low, Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 2012

[5] S. H. Low, Convex relaxation of optimal power flow. *IEEE Transactions on Control of Network Systems*, 1(1):15–27, 2014.

A dynamic formulation for the crude oil procurement problem

Thomas Martin
CERMICS, Ecole Nationale des Ponts et Chaussées

Joint work with ***M. De Lara, P. Carpentier, J.P. Chancelier*** **and** ***V. Leclère***

Keywords : Stochastic Optimization, Multistage, Risk Measure, Stopping Problem

Crude oil procurement is subject to variations in shipping delays and random costs. The optimal procurement problem aims at buying oil for refineries so as to maximize net revenue. The plants owned by a company have different specifications, and so have the crude oil available on the market. In our setting of the problem, we focus on the stochasticity of the costs of oil, shipping and of the prices of the products. The dynamic of the system over time, the shipping delays, and the running of a refinery are supposed to be deterministic.
In this presentation, we will first focus on identifying controls and states variables to build a suitable model for the crude oil procurement problem. The resulting problem is a multistage stochastic optimization problem, which, due to its huge size, is impossible to solve as such. We then discuss the model and explain why the problem can be considered as a multi-stopping time problem. Finally, we take advantage of the peculiar structure of the problem to reduce its size and discuss possible resolution techniques.

Finite State Mean Field Games: Common Noise and Selection of Equilibria

François Delarue
Université Côte d'Azur

Joint work with *Alekos Cecchin*

Keywords : mean field games, potential games, common noise, selection

We address restoration of uniqueness for finite state MFGs by using a relevant form of common noise and then show, in the potential case, how this may help for selecting solutions to the corresponding MFG without common noise.

Discrete-time MFGs with risk-averse agents

Pierre Lavigne
CMAP-INRIA

Joint work with *J. Frédéric Bonnans, Laurent Pfeiffer*

Keywords : Mean field games, risk measure, stochastic control

Classical mean field game theory [6, 4, 5] focusses on risk neutral agents. However in many economics applications it is very natural to consider risk aversion. Risk sensitive case and worst case have been studied by T. Başar and al. in recent works [2], [3]. Our approach is different, we use the theory of risk measure developed by Artzner and al. [1] to build an economic framework in which agents interacts through the prices under risk aversion. We study a discrete time, continuous space system of dynamic programming and Kolmogorov equations. One main feature of the dynamic programming equation is the presence of a risk measure. This equation can be derived from a general cost designed with a composite risk measure [7, 8]. The study of the dynamic programming equation is performed via the theory of proximal operators. We show the existence of an equilibrium of the game via a classical fixed point technique, under suitable assumptions.

References

[1] Philippe Artzner and Freddy Delbaen and Jean-Marc Eber and David Heath, *Coherent Measures of Risk*,Mathematical Finance, 3 (1999), pp. 203-228.

[2] Tamer Basar and Hamidou Tembine and Quanyan Zhu, *Risk-Sensitive Mean Field Games*, IEEE Transactions on Automatic Control, 4 (2014), pp. 835-850.

[3] Tamer Basar and Jun Moon, *Linear Quadratic Risk-Sensitive and Robust Mean Field Games*, IEEE Transactions on Automatic Control, 3 (2017), pp. 1062-1077.

[4] Peter E. Caines and Minyi Huang and Roland P. Malhamé, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Communications in Information & Systems, 3 (2006), 221-252

[5] Pierre Cardaliaguet and Charles-Albert Lehalle, *Mean Field Game of Controls and An Application To Trade Crowding*, arXiv preprint arXiv:1610.09904, (2016)

[6] Jean-Michel Lasry and Pierre-Louis Lions, *Mean field games*, Japanese journal of mathematics, 2 (2007), pp. 229-260.

[7] Andrzej Ruszczyński and Alexander Shapiro, *Conditional Risk Mappings*, Mathematics of Operations Research, 3 (2006), 544-261

[8] Andrzej Ruszczyński, *Risk-averse Dynamic Programming for Markov Decision Processes*, Mathematical Programming, 2 (2010), 235-261

Primal-Dual algorithm for Lambertian Shape from Shading

Hamza Ennaji
Limoges University

Joint work with *Noureddine Igbida, Nguyen Van Thanh*

Keywords : Hamilton-Jacobi equations, Optimization, Shape from Shading

Shape from Shading is a classical problem in compute vision. It consists in reconstructing the 3D shape of an object from a 2D image. A PDE formulation of this problem gives rise to a PDE of the form $H(x, Du) = 0$ on Ω where Ω is a domain of $\mathbb{R}^d$ and $H : \Omega \times \mathbb{R}^d \to \mathbb{R}$ is the Hamiltonian. To every Hamiltonian one can associate a Finsler metric σ which characterise the set of all subsolutions of HJ equation via its dual σ^*. We show [1] that the maximal viscosity subsolution is also solution of a maximization problem of the form

$$\max_u \{ \int_\Omega u dx : \ \sigma^*(x, \nabla u) \leq 1 \ + \ \text{boundary conditions.} \}$$

Using duality, we end up with a problem of the form

$$\inf_u \sup_\phi \mathcal{F}(u) + \langle \nabla u, \phi \rangle - \mathcal{G}^*(\phi) \tag{1}$$

for appropriate functionals $\mathcal{F}$ and $\mathcal{G}$. The problem (1) falls into the scope of [3] and thus offers an alternative to the commonly used approaches to tackle the SfS problem.

References

[1] N.Igbida, H.Ennaji, V.T.Nguyen, *Augmented Lagrangian Method for HJ equations*, 2020.

[2] N.Igbida, H.Ennaji, V.T.Nguyen, *Primal-Dual algorithm for Lambertian Shape from Shading*, 2020, in preparation

[3] Chambolle, Antonin and Pock, Thomas, *A first-order primal-dual algorithm for convex problems with applications to imaging*, 2011.

Nonconvex Quadratic Optimization with Gurobi

Robert Luce
Gurobi

Keywords : Nonconvex optimization, branch-and-bound algorithm, quadratic optimization

We discuss the solution of nonconvex mixed integer quadratic optimization problems. The approach that Gurobi takes on solving such problems to global optimality is based on the McCormick relaxation of bilinear constraints: Each quadratic constraint is reformulated as a set of bilinear constraints, and their McCormick relaxations are then used within a branch-and-bound algorithm. We outline the key algorithmic components that are involved: Reformulations, branching rules and cutting planes.