Modern Techniques of Very Large Scale Optimization

Book of Abstracts

19th-20th May 2022, Edinburgh

THE UNIVERSITY
of EDINBURGH

Organizing Committee

- **Cipolla Stefano** (The University of Edinburgh, School of Mathematics)
- **Gondzio Jacek** (The University of Edinburgh, School of Mathematics)
- **Zanetti Filippo** (The University of Edinburgh, School of Mathematics)

Sponsors

Google

Fast $(1 + x)$-order methods for linear programming problems (2020-2021)

Contents

Modern Techniques of Very Large Scale Optimization

Schedule

The program below refers to the UK time zone (GMT+1)

May 19th, 2022	
09.00 ⟶ 09.20 **Opening Remarks**	
09.20 ⟶ 10.45 Chair: J. Gondzio	
09.20 ⟶ 10.20	J. Eckstein
10.20 ⟶ 10.45	A. Sola Vilalta
10.45 ⟶ 11.20 Coffee Break	
11.20 ⟶ 12.10 Chair: S. Cipolla	
11.20 ⟶ 11.45	M. Kocvara
11.45 ⟶ 12.10	S. Zhao
12.10 ⟶ 13.45 Lunch	
13.45 ⟶ 17.00 Chair: J. Eckstein	
13.45 ⟶ 14.35	H. Wolkowicz (online)
14.35 ⟶ 15.25	S. Bellavia
15.25 ⟶ 16.00 Coffee Break	
16.00 ⟶ 17.00	Y. Ye (online)
19.00 Social Dinner: The Rabbit Hole	

May 20th, 2022	
09.30 ⟶ 10.45 Chair: D. di Serafino	
09.30 ⟶ 10.20	R. Chan (online)
10.20 ⟶ 10.45	F. Zanetti
10.45 ⟶ 11.20 Coffee Break	
11.20 ⟶ 12.35 Chair: E. Bednarczuk	
11.20 ⟶ 12.10	M. Figueiredo
12.10 ⟶ 12.35	G. Scrivanti
12.35 ⟶ 14.10 Lunch	
14.10 ⟶ 15.25 Chair: M. Figueiredo	
14.10 ⟶ 15.00	D. di Serafino
15.00 ⟶ 15.25	S. Cipolla
15.25 ⟶ 16.00 Coffee Break	
16.00 ⟶ 17.15 Chair: S. Bellavia	
16.00 ⟶ 16.50	E. Bednarczuk
16.50 ⟶ 17.15	G. Bruccola
17.15 ⟶ 17.20 Final Remarks	

Abstracts

On dynamical system related to a primal-dual scheme for finding zeros of the sum of maximally monotone operators

Ewa Bednarczuk[1,2], Raj Narayan Dhara[1,3] and Krzysztof Rutkowski[1,4]

[1] Systems Research Institute, PAS

[2] Warsaw University of Technology. ewa.bednarczuk@ibspan.waw.pl

[3] Department of Mathematics and Statistics, Masaryk University

[4] Cardinal Stefan Wyszynski University

In the talk we introduce a dynamical system to the problem of finding zeros of the sum of two maximally monotone operators. The trajectories of the proposed dynamical system converge strongly to a primal-dual solution of the considered problem. Under explicit time discretization of the dynamical system we obtain the best approximation algorithm for solving coupled monotone inclusion problem.

Let $\mathcal{H}$, $\mathcal{G}$ be Hilbert spaces. We consider the problem of finding $p \in \mathcal{H}$ such that

$$0 \in Ap + L^*BLp, \tag{P}$$

where $A : \mathcal{H} \to \mathcal{H},\ B :\ \mathcal{G} \to \mathcal{G}$ are maximally monotone operators, $L : \mathcal{H} \to \mathcal{G}$ is a bounded, linear operator. Together with problem (P) we consider the dual problem of finding $v^* \in \mathcal{G}$ such that

$$0 \in -LA^{-1}(-L^*v^*) + B^{-1}v^*. \tag{D}$$

To problems (P) and (D) we associate Kuhn-Tucker set defined as

$$Z := \{(p, v^*) \in \mathcal{H} \times \mathcal{G} \mid -L^*v^* \in Ap \quad \text{and} \quad Lp \in B^{-1}v^*\}. \tag{Z}$$

We propose the following dynamical system, solution of which asymptotically approaches solution of (P)-(D),

$$\begin{aligned} &\dot{x}(t) = Q(\bar{w}, x(t), \mathbb{T}x(t)) - x(t), \quad t \geq 0,\\ &x(0) = x_0, \end{aligned} \tag{S}$$

where $x_0, \bar{w} \in \mathcal{H} \times \mathcal{G}$, $\mathbb{T} :\ \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G}$, fixed point set of the operator $\mathbb{T}$ is Z ($\operatorname{Fix} \mathbb{T} = Z$), with Z defined by (Z) and $Q : (\mathcal{H} \times \mathcal{G})^3 \to \mathcal{H} \times \mathcal{G}$,

$$Q(\bar{w}, b, c) := P_{H(\bar{w},b)\cap H(b,c)}(\bar{w}), \tag{1}$$

is the projection P of the element $\bar{w}$ onto the set $H(\bar{w}, b) \cap H(b, c)$ which is the intersection of two hyperplanes of the form

$$H(z_1, z_2) := \{h \in \mathcal{H} \times \mathcal{G} \mid \langle h - z_2 \mid z_1 - z_2\rangle \leq 0\}, \quad z_1, z_2 \in \mathcal{H} \times \mathcal{G}. \tag{2}$$

In particular, $H(\bar{w}, b) = \{h \in \mathcal{H} \times \mathcal{G} \mid \langle h - b \mid \bar{w} - b\rangle \leq 0\}$.

The most essential difference between (S) and the systems investigated in the literature is that, in general, one cannot expect that the vector field Q appearing in (S) is globally Lipschitz with respect to variable x.

Under explicit discretization with step size equal to one the system (S) becomes the best approximation algorithm for finding fixed point of $\mathbb{T}$

$$x_{n+1} = Q(\bar{w}, x_n, x_{n+1/2}), \quad n \in \mathbb{N}. \tag{3}$$

with the choice of $x_{n+1/2} := \mathbb{T}(x_n)$ and the starting point x_0. The characteristic feature of this algorithm is the strong convergence of the sequence x_n to a fixed point of $\mathbb{T}$.

When $A = \partial f$, $B = \partial g$, $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, $g : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ are proper convex, lower semicontinuous (l.s.c.) functions, the problem (P) (if solvable) reduces to finding a point $p \in \mathcal{H}$ solving the following minimization problem

$$\text{minimize}_{p\in\mathcal{H}} f(p) + g(Lp) \tag{4}$$

and (D) reduces to finding a point $v^* \in \mathcal{G}$ solving the following maximization problem

$$\text{maximize}_{v^*\in\mathcal{G}} - f^*(-L^*v^*) - g^*(v^*). \tag{5}$$

Optimization Methods based on Random Models and Examples from Machine Learning

Stefania Bellavia[1]

[1] University of Florence. stefania.bellavia@unifi.it

We discuss trust region methods using random models and stochastic function estimates. The primary motivation for developing these methods is the need to solve optimization problems that arise in machine learning, which because of the enormous amounts of data involved in each computation of the function and gradient, usually require a stochastic optimization approach. The main issues that we address in this talk are the adaptive choice of step-size and of the the accuracy conditions that the models and function values have to satisfy. In particular we assume that our models satisfy some good quality conditions with some probability fixed, but can be arbitrarily bad otherwise. Both theoretical and computational results will be presented.

Finding global solutions for a class of possibly nonconvex QCQP problems through the S-lemma

Giovanni Bruccola[1]

[1] Systems Research Institute, Polish Academy of Sciences, ITN Marie-Curie Project TraDe-Opt Giovanni.Bruccola@ibspan.waw.pl

In this talk we provide necessary and sufficient (KKT) conditions for global optimality for a new class of possibly nonconvex quadratically constrained quadratic programming (QCQP) problems, denoted by S-QCQP. The proof relies on a generalized version of the S-Lemma, stated in the context of general QCQP problems. Moreover, we prove the exactness of the SDP and the SOCP relaxations for S-QCQP.

Based on the join work with Ewa M. Bednarczuk (Systems Research Institute, Polish Academy of Sciences).

A 3-stage Spectral-spatial Method for Hyperspectral Image Classification

Raymond H. Chan[1] and Ruoning Li[1]

[1] Department of Mathematics, City University of Hong Kong. raymond.chan@cityu.edu.hk

Hyperspectral images often have hundreds of spectral bands of different wavelengths captured by aircraft or satellites that record land coverage. Identifying detailed classes of pixels becomes feasible due to the enhancement in spectral and spatial resolution of hyperspectral images. In this work, we propose a novel framework that utilizes both spatial and spectral information for classifying pixels in hyperspectral images. The method consists of three stages. In the first stage, the pre-processing stage, the Nested Sliding Window algorithm is used to reconstruct the original data by enhancing the consistency of neighboring pixels, and then Principal Component Analysis is used to reduce the dimension of data. In the second stage, Support Vector Machines are trained to estimate the pixel-wise probability map of each class using the spectral information from the images. Finally, a smoothed total variation model is applied to smooth the class probability vectors by ensuring spatial connectivity in the images. We demonstrate the superiority of our method against three state-of-the-art algorithms on six benchmark hyperspectral data sets with 10 to 50 training labels for each class. The results show that our method gives the overall best performance in accuracy. Especially, our gain in accuracy increases when the number of labeled pixels decreases. Therefore our method is of great practical significance since expert annotations are often expensive and difficult to collect.

Primal Dual Regularized IPM: a Proximal Point perspective

Stefano Cipolla[1], Jacek Gondzio[1]

[1] School of Mathematics, The University of Edinburgh. scipolla@exseed.ed.ac.uk

Computational evidence suggests that the Primal-Dual Regularization for Interior Point Methods (IPMs) is a successful technique able to stabilize and to speed-up the linear algebra used in IPM implementations [1]. On the other hand, many issues remain open when IPMs are used in their primal-dual regularized form and, in particular, to the best of our knowledge, the known convergence theory requires strong assumptions on the uniform boundedness of the Newton directions [2]. Recently, the study of the interaction of primal-dual regularized IPMs with the Augmented Lagrangian Method and the Proximal Point Algorithm has permitted to prove the convergence when the regularization parameter is driven to zero at a suitable speed [3].

In this talk, we will show that it is possible to naturally frame the primal-dual regularized IPMs in the context of the Proximal Point Algorithm [4]. Among the benefits of the proposed approach, we will show how convergence can be guaranteed without any supplementary assumptions and how the rate of convergence can be explicitly estimated in relation to (fixed) regularization parameter. Additionally, numerical results proving the efficiency and reliability of the proposed approach will be presented.

References

[1] Altman, A., Gondzio, J. (1999) Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization, Optimization Methods and Software, 11:1-4, 275-302.

[2] Friedlander, M. P., Orban D. (2012) A primal-dual regularized interior-point method for convex quadratic programs, Mathematical Programming Computation 4.1 : 71-107.

[3] Pougkakiotis, S., Gondzio J. (2021) An interior point-proximal method of multipliers for convex quadratic programming, Computational Optimization and Applications 78.2 : 307-351.

[4] Cipolla, S., Gondzio J. (2022) Proximal stabilized Interior Point Methods for quadratic programming and low-frequency-updates preconditioning techniques, https://doi.org/10.48550/arXiv.2205.01775

Efficient Solution of Sparse Optimization Problems via Interior Point Methods

Daniela di Serafino[1], Valentina De Simone[2], Marco Viola[2], Jacek Gondzio,[3] and Spyros Pougkakiotis,[4]

[1] University of Naples Federico II. daniela.diserafino@unina.it

[2] University of Campania "L. Vanvitelli"

[3] School of Mathematics, University of Edinburgh

[4] Yale University

We are concerned with the efficient solution of large optimization problems that are expected to yield sparse solutions. Problems of this type arise in many application areas (e.g., portfolio optimization, signal and image processing, classification in statistics and machine learning, inverse problems and compressed sensing) and are often solved by specialized first-order methods, although they often struggle with not-so-well conditioned problems. For these problems we developed variants of an Interior Point-Proximal Method of Multipliers that use suitable linear algebra solvers and take advantage of the expected sparsity in the optimal solution [1]. Results of computational experiments on a variety of problems, including comparisons with state-of-the art methods, show that Interior Point Methods can provide a significant advantage over first-order approaches.

References

[1] V. De Simone, D. di Serafino, J. Gondzio, S. Pougkakiotis, M. Viola, *Sparse Approximations with Interior Point Methods*, to appear in SIAM Review (2022).

The ADMM: Past, Present, and Future

Jonathan Eckstein[1]

[1] Rutgers University. jeckstei@business.rutgers.edu

Over the past 15 years, the alternating direction method of multipliers (ADMM) has become a standard optimization method. This talk will cover the origins of the ADMM, its subsequent development, and what to expect in the future. The origins of the ADMM are somewhat unusual in that it was discovered computationally before it was analyzed. Its convergence analysis is also noteworthy because, while the ADMM may outwardly appear to be a dual ascent method, the natural analyses center on reducing the distance to certain fixed points combining primal and dual variables. The nature of these analyses explains the difficulty of proving convergence of natural variants of the algorithm that change the penalty parameter between iterations or involve sums of more than two functions. We will also cover some currently known variations on the ADMM and the problem formulation features that tend to distinguish between successful and unsuccessful applications. Finally, the talk will briefly address what we may expect for the future and what other operator splitting methods might become viable members of the optimization toolbox.

ADMM in Imaging Inverse Problems: Review of a Line of Work

Mário A. T. Figueiredo[1]

[1] Instituto de Telecomunicações. Instituto Superior Técnico, University of Lisbon. mario.figueiredo@lx.it.pt

I will review a line of work on the use of the ADMM (alternating direction method of multipliers) that we have pursued to address a wide variety of imaging inverse problems. At the core of this line of work is a way of using ADMM to tackle optimization problems where the objective function is the sum of two or more convex functions, each of which having an efficiently computable proximity operator and each possibly composed with a linear operator. The approach is illustrated on a variety of problems, namely: image restoration and reconstruction from linear observations (e.g., compressive sensing, image deblurring, inpainting), which may be contaminated with Gaussian or Poisson noise, using synthesis, analysis, or hybrid regularization, and unconstrained or constrained optimization formulations. In the second part, I will address the so-called plug-and-play (PnP) approach, wherein a formal regularizer is replaced with a black-box denoiser, aiming at leveraging state-of-the-art denoisers in more general inverse problems. Since these denoisers lack an explicit optimization formulation, classical results on the convergence of ADMM cannot be directly invoked.

An interior-point method for Lasserre relaxations of unconstrained binary quadratic optimization problems

Michal Kocvara[1]

[1] School of Mathematics, University of Birmingham. m.kocvara@bham.ac.uk

The aim of this paper is to solve linear semidefinite programs arising from Lasserre relaxations of unconstrained binary quadratic optimization problems. For this we use an interior point method with a preconditioned conjugate gradient method solving the linear systems. The preconditioner utilizes the low-rank structure of the solution of the relaxations. In order to fully utilize this, we re-write the moment relaxations. To treat the arising linear equality constraints we use an ℓ_1-penalty approach within the interior-point solver. The efficiency is demonstrated by numerical experiments and comparison with a state of the art semidefinite solver.

A Convex-Non-Convex approach for Directional Total Variation

Gabriele Scrivanti[1], Emilie Chouzenoux[1] and Jean-Christophe Pesquet[1]

[1] Université Paris-Saclay, Inria, CentraleSupélec, CVN, Gif-sur-Yvette. gabriele.scrivanti@centralesupelec.fr

A wide class of problems arising in image and signal processing is represented by sparsity- regularised variational models. The most natural sparsity inducing penalty is the l0-pseudo-norm, but it forces the related problem to be NP hard and nonconvex. A popular convex variational surrogate is the l1-norm, though it has the drawback of under-estimating the high amplitude components of the considered signal. Nonconvex variational regularisers manage to overcome this issue, at the cost of introducing suboptimal local minima in the objective function. An efficient solution to keep only the best traits of these regularisers is represented by Convex Non Convex strategies: they consist of building convex objective functionals that include nonconvex regularisation terms. Suitable CNC strategies have been designed for more and more general classes of problems. A CNC form of Total Variation has shown to get around the well known problems of boundary reduction and staircasing effect related to the classic TV regularisation for image Restoration and Segmentation. We propose a new variational formulation extending the CNC strategies to the recent Directional Total Variation model. We identify a Primal Dual procedure to efficiently address the resulting optimisation problems and provide experimental results that support the use of the presented regularisation method.

ADMM-based Unit and Time Decomposition for Price Arbitrage by Cooperative Price-Maker Electricity Storage Units

Albert Sola Vilalta[1]

[1] School of Mathematics, The University of Edinburgh albert.sola@ed.ac.uk

Decarbonization via the integration of renewables poses significant challenges for electric power systems, but also creates new market opportunities. Electric energy storage can take advantage of these opportunities while providing flexibility to power systems that can help address these challenges. We propose a solution method for the optimal control of multiple price-maker electric energy storage units that cooperate to maximize their total profit from price arbitrage. The proposed method can tackle the nonlinearity introduced by the price-maker assumption. The main novelty of the proposed method is the combination of a decomposition by unit and a decomposition in time. The decomposition by unit is based on the Alternating Direction Method of Multipliers (ADMM) and breaks the problem into several one-unit subproblems. Every subproblem is solved using an efficient algorithm for one-unit problems from the literature that exploits an on the fly decomposition in time, and this results in a time decomposition for the whole solution method. Our numerical experiments show very promising performance in terms of accuracy and computational time. In particular, they suggest that computational time scales linearly with the number of subproblems, i.e., storage units.

A Restricted Dual Peaceman-Rachford Splitting Method for a Strengthened DNN Relaxation for QAP

Henry Wolkowicz[1]

[1] University of Waterloo. hwolkowi@uwaterloo.ca

Splitting methods in optimization arise when one can divide an optimization problem into two or more simpler subproblems. They have proven particularly successful for relaxations of problems involving discrete variables. We revisit and strengthen splitting methods for solving doubly nonnegative, DNN, relaxations of the particularly difficult, NP-hard quadratic assignment problem, QAP. We use a modified restricted contractive splitting method, PRSM, approach. In particular, we show how to exploit redundant constraints in the subproblems. Our strengthened bounds exploit these new subproblems, as well as new dual multiplier estimates, to improve on the bounds and convergence results in the literature.

Randomized ADMM and its applications

Yinyu Ye[1]

[1] Department of Management Science and Engineering and Institute for Computational and Mathematical Engineering, Stanford University. yyye@stanford.edu

Randomization strategy has been widely used in designing optimization algorithms to achieve properties that deterministic algorithms cannot do, such as SGD, BCD, etc. In this talk, we show how randomized techniques help in greatly improving the efficiency of the multi-block alternating direction method with multipliers (MBADMM). We introduce few randomized strategies for improving the efficiency of the method. First, we present benefit of data exchange in distributed optimization and learning. Theoretically, we show that there exists a class of data structure such that data exchange would improve the convergence speed. We then provide numerical evidence on showing that with small amount (around 5%) of data sharing among different blocks would hugely benefit the efficiency of optimization algorithm in large scale learning models. We further demonstrate data-sharing could also benefit other optimization and estimation algorithms, balancing the algorithmic efficiency and data privacy. Secondly, we present benefits of random exchange of variables among the blocks for quadratic mixed integer programming. Specifically, we present two techniques: one utilizing heuristics to escape local minima and the other using the branch-and-bound methodology. Computational studies are presented to show that the method either find a better-quality solution in a limited time or reduce the number of the branch-and-bound nodes.

Applications of interior point method for very large problems arising in imaging and optimal transport

Filippo Zanetti[1], Jacek Gondzio[1], Samuli Siltanen[2], Salla Latva-Aijo[2] and Matti Lassas[2]

[1] School of Mathematics, The University of Edinburgh. f.zanetti@sms.ed.ac.uk

[2] Department of Mathematics and Statistics, University of Helsinki.

In this talk we discuss how interior point methods (IPMs) can be applied to large scale problems and how they perform with respect to first order methods; in particular, two applications are considered. The first one is an IPM for tomographic imaging with a novel regularization [1] that tries to separate the two materials in the object; the quadratic term is large and dense and can only be accessed through matrix-vector products; we present theoretical results in terms of choice of the regularization parameters and eigenvalue bounds for the preconditioned matrix, as well as experimental data. The second application is a sparse IPM for discrete optimal transport problems, where the constraint matrix is of huge dimension but highly sparse and structured; we present a theoretical result about the sparsity structure of the optimal solution that allows us to mix iterative and direct solvers efficiently, as well as experimental results.

References

[1] J. Gondzio, M. Lassas, S. Latva-Aijo, S. Siltanen and F. Zanetti *Material-separating regularizer for multi-energy x-ray tomography*, Inverse Problems, 38 (2022).

Strong SDP bounds for large k-equipartition problem via augmented Lagrangian method combined with projection

Shudian Zhao[1]

[1] Alpen-Adria-Universität Klagenfurt shudian.zhao@aau.at

This presentation investigates the quality of doubly nonnegative (DNN) relaxation, i.e., relaxations having matrix variables that are both positive semidefinite and nonnegative, strengthened by polyhedral cuts the k-equipartition problem. This problem is to cluster the vertex set of a graph equally into k disjoint subsets such that the sum of weights of edges joining different sets is minimized. After reducing the size of the relaxation by facial reduction, we solve it by a cutting-plane algorithm that combines an augmented Lagrangian method with Dykstra's projection algorithm. We are the first to show the power of DNN relaxations with additional cutting planes for the k-equipartition on large benchmark instances up to 1000 vertices. Computational results show impressive improvements in strengthened DNN bounds. Since many components of our algorithm are general, our algorithm is suitable for solving various DNN relaxations with a large number of cutting planes.

Author Index