Book of Abstracts

Algorithmic Optimization: Tools for AI and Data Science
(ALGOPT2024)

https://sites.uclouvain.be/algopt2024/

Université catholique de Louvain

August 27-30, 2024

Last updated on August 26, 2024

Introduction

Optimization methods are fundamental to data science and AI, enabling critical tasks such as sparse data recovery, clustering, dimensionality reduction, and data completion. They are also essential for training machine learning models, including neural networks. However, existing optimization methods are being constantly challenged by the particular features of modern applications. For example, emerging problems related to data science and AI often involve a huge number of variables, complex data structure, and may require distributed data support. Consequently, there is a rising demand for the development of new optimization methods customized for specific applications and ideally supported by provable efficiency guarantees. This workshop aims to bring together experts interested in the theory, design, and application of optimization methods. It will serve as a platform for sharing the latest advancements in Algorithmic Optimization, promoting idea exchange and scientific collaborations.

The workshop will also honor Professor Yurii Nesterov,
celebrating 50 years of his research career in Optimization.

Organizing and Scientific Committee

Organizing Committee

- François Glineur
- Geovani Nunes Grapiglia
- Nancy Guillaume
- Séverine De Visscher

Scientific Committee

- Yurii Nesterov
- François Glineur
- Geovani Nunes Grapiglia

Sponsors

- European Research Council
- UCLouvain (CORE and LIDAM institutes, ICTEAM institute)

Invited Speakers

Masoud Ahookhosh (University of Antwerp, Belgium)

Forward-backward envelope meets arbitrary-order regularization: Framework and algorithms

In this talk, we introduce a version of the forward-backward envelope with arbitrary-order regularization (HiFBE), which is a smooth approximation for nonsmooth and nonconvex functions. This leads to a natural generalization of the classical forward-backward envelope with quadratic regularization. After studying the fundamental properties of HiFBE, we investigate its differentiability under the prox-regularity of the composite objective function. Afterward, we introduce a boosted forward-backward splitting algorithm (Boosted FBS), which combines the arbitrary-order forward-backward splitting with a derivative-free line search. While the well-definedness and the subsequential convergence are studied under mild assumption, the global and linear convergence are investigated under extra KL inequality. Our preliminary numerical experiments indicate a promising behavior of the presented algorithm, which validates our theoretical foundation.

Francis Bach (INRIA, France)

Geometry-dependent matching pursuit: a transition phase for convergence on linear regression and LASSO

Greedy first-order methods, such as coordinate descent with Gauss-Southwell rule or matching pursuit, have become popular in optimization due to their natural tendency to propose sparse solutions and their refined convergence guarantees. In this work, we propose a principled approach to generating (regularized) matching pursuit algorithms adapted to the geometry of the problem at hand, as well as their convergence guarantees. Building on these results, we derive approximate convergence guarantees and describe a transition phenomenon in the convergence of (regularized) matching pursuit from underparametrized to overparametrized models (joint work with Céline Moucer and Adrien Taylor).

Nicolas Boumal (EPFL, Switzerland)

No acceleration for convex optimization on negatively curved spaces

I will talk about work with Christopher Criscitiello: Hamilton and Moitra (2021) showed that, in certain regimes, it is not possible to accelerate Riemannian gradient descent in the hyperbolic plane if we restrict ourselves to algorithms which make queries in a (large) bounded domain and which receive gradients and function values corrupted by a (small) amount of noise. We show that acceleration remains unachievable for any deterministic algorithm which receives exact gradient and function-value information (unbounded queries, no noise). Our results hold for the classes of strongly and nonstrongly geodesically convex functions, and for a large class of Hadamard manifolds.

Coralia Cartis (University of Oxford, UK)

Tensor methods for nonconvex optimization

We investigate the use of high-order polynomial local models within adaptive regularization frameworks for solving nonconvex optimization problems. Though tensor-based models and methods have preoccupied optimization researchers in the past, this topic is experiencing a recent resurgence, possibly originating with the development of carefully-regularized variants of Newton's method. It was shown, in both the convex and nonconvex worlds, that regularization algorithms allow higher-than-second derivatives to be naturally incorporated into their local model construction, while controlling convergence by means of a regularization parameter present in these models. Furthermore, the global rate of convergence/worst-case evaluation complexity of the ensuing methods proved to be optimal within their respective algorithmic classes that use corresponding derivative orders or their approximations. We review these developments, as well as establish connections to polynomial optimization and related techniques, in our efforts to efficiently minimise polynomial subproblems, to local or global optimality. A robust adaptive procedure for the choice of the regularization parameter is key to numerical performance, and we detail some of the intricacies involved in its design. Finally, we bring these ingredients together, in our quest to develop practical tensor methods, with exact and inexact derivatives, subproblem solutions and adaptive parameters, and ask the challenging question: are tensor methods preferable to second-order ones in numerical practice?

Nikita Doikov (EPFL, Switzerland)

Spectral Preconditioning for Gradient Methods on Graded Non-convex Functions

The performance of optimization methods is often tied to the spectrum of the objective Hessian. Yet, conventional assumptions, such as smoothness, do often not enable us to make finely-grained convergence statements – particularly not for non-convex problems. Striving for a more intricate characterization of complexity, we introduce a unique concept termed graded non-convexity. This allows to partition the class of non-convex problems into a nested chain of subclasses. Interestingly, many traditional non-convex objectives, including partially convex problems, matrix factorizations, and neural networks, fall within these subclasses. Then, we propose gradient methods with spectral preconditioning, which employ inexact top eigenvectors of the Hessian to address the ill-conditioning of the problem, contingent on the grade. Our analysis reveals that these new methods provide provably superior convergence rates compared to basic gradient descent on applicable problem classes, particularly when large gaps exist between the top eigenvalues of the Hessian.

Joint work with Sebastian U. Stich and Martin Jaggi.

Paul Van Dooren (Université catholique de Louvain)

Assigning Stationary Distributions to Sparse Stochastic Matrices

The target stationary distribution problem (TSDP) is the following: given an irreducible stochastic matrix G and a target stationary distribution $\hat{\mu}$, construct a minimum norm perturbation, Δ, such that $\hat{G} = G + \Delta$ is also stochastic and has the prescribed target stationary distribution, $\hat{\mu}$. In this paper, we revisit the TSDP under a constraint on the support of Δ, that is, on the set of non-zero entries of Δ. This is particularly meaningful in practice since one cannot typically modify all entries of G. We first show how to construct a feasible solution $\hat{G}$ that has essentially the same support as the matrix G. Then we show how to compute globally optimal and sparse solutions using the componentwise ℓ_1 norm and linear optimization. We propose an efficient implementation that relies on a column-generation approach which allows us to solve sparse problems of size up to $10^5 \times 10^5$ in a few minutes.

Radu-Alexandru Dragomir (Télécom Paris, France)

Gradient methods for quartic problems

Many tasks in signal processing and machine learning involve minimizing polynomials of degree four. These include phase retrieval, matrix factorization, sensor network localization and many more. In this talk, I will give an overview of the challenges of quartic minimization as well as some complexity results. In particular, we will focus on a particular class of convex quartic problems, and we analyze the notion of quartic condition number. We design an algorithm for reducing this condition number. To do so, we build a preconditioner using a generalized version of the Lewis weights (a.k.a leverage scores), and we show that it is optimal in some specific sense.

Pavel Dvurechensky (WIAS, Germany)

On some results on minimization involving self-concordant functions and barriers

In this talk, I will give an overview of some recent results on minimization problems that involve self-concordant functions (SCF) or barriers (SCB). In the first part, I will focus on non-convex constrained minimization problems where the feasible set is given as the intersection of an affine subspace and either a convex cone or a general convex set. In the first case, we use logarithmically-homogeneous SCBs for the cone. In the second case, we use SCBs for the convex set. We propose first- and second-order algorithms that allow us to find a suitably defined class of approximate first- or second-order KKT points with the worst-case iteration complexity similar to unconstrained problems, namely $\mathcal{O}(\epsilon^{-2})$ (first-order) and $\mathcal{O}(\epsilon^{-3/2})$ (second-order), respectively. In the second part, I will focus on minimization of SCFs without the barrier property. In particular, we show that path-following methods for unconstrained SCF minimization have superlinear convergence and better complexity guarantees than those of the classical Damped Newton Method. If time allows, I will discuss constrained minimization of SCFs by Frank-Wolfe algorithms, where we obtain classical $\mathcal{O}(\epsilon^{-1})$ complexity and linear convergent methods by advanced versions of the Frank-Wolfe algorithms.

Mihai Iulian Florea (Université catholique de Louvain, Belgium)

Gradient Methods with Memory

The maintenance of a bundle, with the associated oracle-based piece-wise linear model of the objective, is a standard and powerful approach in nonsmooth optimization but seldom applied to smooth objectives due to perceived computational difficulties. In this walk we show that employing the bundle can lead to sometimes spectacular performance gains in smooth optimization under very relaxed assumptions. For instance, combining the bundle with line-search produces a Gradient Method with Memory that can take steps several orders of magnitude larger than the Gradient Method with line-search. Exact Gradient Methods with Memory render this approach parameter-free and further benefit from acceleration. We can apply the bundle in the estimate sequence itself to produce an oracle-free convergence guarantee adjustment procedure. The overhead for composite problems can be rendered negligible. On methods recently obtained via Performance Estimation that are absolutely optimal, even up to the proportionality constant, the optimized estimate sequence constitutes the only known adaptive mechanism. Combining the bundle with adaptive restart produces optimal parameter-free methods that are competitive on a variety of signal processing and machine learning problems with state-of-the-art schemes that have full knowledge of the problem parameters. This is joint work with Prof. Yurii Nesterov.

Robert Freund (MIT, USA)

The Role of Level-Set Geometry on the Performance of PDHG for Conic Linear Optimization

We consider solving huge-scale instances of (convex) conic linear optimization problems, at the scale where matrix-factorization-free methods are attractive or necessary. The restarted primal-dual hybrid gradient method (rPDHG) – with heuristic enhancements and GPU implementation – has been very successful in solving huge-scale linear programming (LP) problems. Here we extend the study of rPDHG to encompass Conic Linear Optimization as well. We present a new theoretical analysis of rPDHG for general (convex) conic linear optimization and for LP as a special case thereof. We show a relationship between geometric measures of the primal-dual (sub-)level sets and the convergence rate of rPDHG. We also show how central-path-based linear transformations–including conic rescaling – can markedly enhance the convergence rate of rPDHG. Last of all, we present computational results that demonstrate how rescalings can accelerate convergence to high-accuracy solutions, and lead to more efficient methods for huge-scale LP problems, and that outperform interior-point methods. This is joint work with Zikai Xiong.

Michel Gevers (Université catholique de Louvain, Belgium)

Recent results on the identification of dynamical networks

We consider dynamical networks with a topology of directed graphs, where the nodes are connected by rational transfer functions. The nodes can be subjected to external excitation signals, they can be measured, or both excited and measured. Our work focuses on the identifiability of the transfer functions of the network on the basis of these excitation and/or measurement signals. This problem was first formalized about 12 years ago. Whereas, initially, the results on network identifiability assumed that all nodes were either excited or measured, in 2019 we published the first results for networks where this constraint was relaxed. For the last 5 years, the search has been for sparse Excitation and Measurement Patterns (EMP), i.e. combinations of excitation and measurements that make the network identifiable, preferably with a small (or the smallest) cardinality. The talk will present some important breakthroughs of the last 5 years.

François Glineur (Université catholique de Louvain, Belgium)

Exact convergence rates of the last iterate in subgradient methods

We study the convergence of the last iterate in subgradient methods applied to the minimization of a nonsmooth convex function with bounded subgradients. We first introduce a proof technique that generalizes the standard analysis of subgradient methods. It is based on tracking the distance between the current iterate and a different reference point at each iteration. Using this technique, we obtain the exact worst-case convergence rate for the objective accuracy of the last iterate of the projected subgradient method with either constant step sizes or constant step lengths. Tightness is shown with a worst-case instance matching the established convergence rate. We also derive the value of the optimal constant step size when performing a given number of iterations. Finally, we introduce a new optimal subgradient method that achieves the best possible last-iterate accuracy after a given number of iterations. Its convergence rate matches exactly the lower bound on the performance of any black-box method on the considered problem class. We also show that no single memory-less method can achieve this optimal rate for any number of iterations.

This is a joint work with Moslem Zamani.

Geovani Grapiglia (Université catholique de Louvain, Belgium)

On the Complexity of Quadratic Penalty Function and Augmented Lagrangian Methods for Nonlinear Programming Problems

In this talk, I will discuss about the worst-case complexity of Quadratic Penalty Function and Augmented Lagrangian methods for nonconvex constrained optimization problems. Specifically, under suitable conditions, I will show bounds for the number of outer iterations required by these methods to find ϵ-approximate KKT points. Additionally, for particular problem classes, I will describe upper bounds for the total number of calls of the oracle, when appropriate pth-order unconstrained methods are employed as inner solvers.

Niao He (ETH Zürich, Switzerland)

Stochastic Optimization under Hidden Convexity

In this talk, we consider a broad class of nonconvex stochastic optimization problems that possess a hidden convexity structure, i.e., those admitting a convex reformulation through via some non-linear mapping. Such problems are prevalent in various data-driven applications, such as optimal control, revenue and inventory management, and convex reinforcement learning. However, solving the convex reformulation is often impractical or infeasible. Instead, we focus on addressing the original hidden convex problems directly using stochastic first-order methods. We first examine the simplest projected stochastic (sub-)gradient methods and present the first sample complexity guarantees for achieving global convergence in both smooth and non-smooth settings. We further introduce new algorithms with improved sample complexity results for special cases and demonstrate their efficiency in revenue management applications.

Tibor Illés (Corvinus University of Budapest, Hungary)

Parabolic Target-Space Interior-Point Algorithm for Weighted Monotone Linear Complementarity Problem

In this talk we present a generalization of the result of Nesterov (2008), the primal–dual interior-point algorithm (IPA) for LP problems. Similarly to the paper of Nesterov, our approach is based on the concept of parabolic target space (PTS). The concept of weighted central path (WCP) in the literature of LCPs, first occurs in a paper of Illés, Roos, Terlaky (1997). We introduce a relaxation of WCP and show that the solution set of the relaxed problem is convex. Later an additional (convex) constraint on the duality gap of the MLCP has been added. Finally, we arrived at a convex feasibility problem (CFP) that has original variables of the MLCP, and those related to the relaxation and the additional constraint. The new variables, naturally satisfies an extra condition leading to the observation of a PTS. The use of the PTS allows us to discuss the new IPA for both the weighted and classical MLCP at the same time. The solution of an MLCP reduces to the solution of a sequence of CFPs. The driving forth of our new IPA lies in the structure of CFPs and the assigned self-concordant barrier function of CFP, and its properties.

Joint work with M. E.-Nagy, Yu. Nesterov and P.R. Rigó.

Anatoli Iouditski (Université Grenoble-Alpes, France)

Reduced variance Stochastic Mirror Descent for Large-Scale Sparse Recovery

Our focus is on applications of Stochastic Approximation to statistical estimation. As opposed to the classical problem setting in which the amplitude of noise of the gradient observation is assumed to be uniformly bounded, herein we assume this amplitude to be related to the "sub-optimality" of the approximate solutions delivered by the algorithm. Such problems naturally arise in a variety of applications, in particular, in the well-known generalized linear regression problem in statistics. However, to the best of our knowledge, none of the existing stochastic approximation algorithms for solving this class of problems attain optimality in terms of the dependence on accuracy, problem parameters, and mini-batch size. We show that "appropriately adjusted" long-step SA, achieve the optimal convergence rate in this setting, attaining the optimal iteration and sample complexities simultaneously. We discuss the application of the proposed algorithms to recovery of sparse solutions.

Etienne de Klerk (Tilburg University, Netherlands)

Semidefinite programming and approximation theory

In this talk we give an overview of the interplay between constructive approximation theory and semidefinite programming (SDP). We will look at SDP approaches to multivariate Chebyshev approximation, as well as the construction of approximation kernels using SDP.

George Lan (Georgia Institute of Technology, USA)

Uniform Optimality for Convex and Nonconvex Optimization

The past few years have witnessed growing interest in the development of easily implementable parameter-free first-order methods to facilitate their applications. In this talk, I will discuss some recent progresses on uniformly optimal methods for convex and nonconvex optimization. By uniform optimality, we mean that these algorithms do not require the input of any problem parameters but can still achieve the best possible iteration complexity bounds for solving different classes of optimization problems. We first consider convex optimization problems under different smoothness levels and show that neither such smoothness information nor line search procedures are needed to achieve uniform optimality. We then consider regularity conditions (e.g., strong convexity and lower curvature) that are imposed over a global scope and thus are notoriously more difficult to estimate. By presenting novel methods that can achieve tight complexity bounds to compute solutions with verifiably small (projected) gradients, we show that such regularity information is in fact superfluous for handling strongly convex and nonconvex problems.

(This talk is based on a joint work with Zhichao Jia and Zhe Zhang).

Goran Lesaja (Georgia Southern University, USA)

Local Self-concordance of Barrier Functions Based on Kernel-functions

Many efficient interior-point methods (IPMs) are based on the use of a self-concordant barrier function on the domain of the problem that has to be solved. In the last two decades, several classes of barrier functions based on kernel functions were considered, including self-regular and eligible kernel functions. They lead to the design of efficient kernel-based IPMs with improved theoretical complexity of long-step IPMs. However, kernel-based barrier functions are not self-concordant on the domain of the problem, except for classical logarithmic barrier function, but despite this fact give rise to efficient IPMs. To explain this fact, we introduce a notion of locally self-concordant barrier functions and we introduce a wide class of new barrier functions based on a class of kernel functions called B-kernel functions, which include many of the self-regular and eligible kernel functions, that we prove are locally self-concordant. In many cases, the (local) complexity numbers of the new barrier functions along the central path are better than the complexity number of the logarithmic barrier function by a factor between 0.5 and 1.

Boris Mordukhovich (Wayne State University, USA)

Convergence of Descent Methods under Kurdyka- Lojasiewicz Properties

This talk prtesents the novel convergence analysis of a generic class of descent methods in nonsmooth and nonconvex optimization under several versions of the Kurdyka- Lojasiewicz (KL) property. Along other results, we prove the finite termination of generic algorithms under the KL property with lower exponents. Specifications are given to convergence rates of some particular algorithms including inexact reduced gradient methods and the boosted algorithm in DC programming. It revealed, e.g., that the lower exponent KL property in the DC framework is incompatible with the gradient Lipschitz continuity for the plus function around a local minimizer. On the other hand, we show that the above inconsistency observation may fail if the Lipschitz continuity is replaced by merely the gradient continuity.

Based on joint work with G. C. Bento, T. S. Mota and Yu. Nesterov.

Ion Necoara (University Politehnica Bucharest, Romania)

Regularized higher-order Taylor approximation methods for nonlinear least-squares

We develop a regularized higher-order Taylor based method for solving nonlinear least-squares problems. At each iteration, we replace each smooth component of the objective function by a higher-order Taylor approximation with an appropriate regularization, leading to a regularized higher-order Taylor approximation (RHOTA) algorithm. We derive global convergence guarantees for RHOTA algorithm. In particular, we prove stationary point convergence guarantees for the iterates generated by RHOTA, and leveraging a Kurdyka-Lojasiewicz (KL) type property of the objective function, we derive improved rates depending on the KL parameter. Furthermore, we extend the scope of our investigation to include the behavior and efficacy of RHOTA algorithm in handling systems of nonlinear equations and optimization problems with nonlinear equality constraints deriving new rates under improved constraint qualifications conditions. Numerical simulations on phase retrieval and output feedback control problems also demonstrate the efficacy and performance of the proposed methods when compared to some state-of-the-art optimization methods and software.

Arkadi Nemirovski (Georgia Institute of Technology, USA)

Some Applications of Convex Optimization in Control and Statistics

In this talk, joint with Anatoli Juditsky, we present and analyze several computation-friendly convex optimization models for several applications coming from Control (synthesis of linear controllers with near-optimal peak-to-peak gain, system identification under deterministic errors in observations of states and inputs) and Statistics (signal recovery in Gaussian observation model with sensing matrix affected by random or uncertain-but-bounded perturbations).

Yurii Nesterov (Université catholique de Louvain, Belgium)

Optimization in the universe of Artificial Intelligence

We discuss new challenges in the modern Science, created by Artificial Intelligence (AI). Indeed, AI requires a system of new sciences, mainly based on computational models. Its development has already started by the progress in Computational Mathematics. In this new reality, Optimization plays an important role, helping the other fields with finding tractable models and efficient methods, and significantly increasing their predictive power.

Panos Patrinos (KU Leuven, Belgium)

Adaptive linesearch-free proximal gradient methods under Hölder gradient continuity

We show that adaptive proximal gradient methods for convex problems are not restricted to traditional Lipschitzian assumptions. Our analysis reveals that a class of linesearch-free methods is still convergent under mere local Hölder gradient continuity, covering in particular continuously differentiable semi-algebraic functions. To mitigate the lack of local Lipschitz continuity, popular approaches revolve around ϵ-oracles and/or linesearch procedures. In contrast, we exploit plain Hölder inequalities not entailing any approximation, all while retaining the linesearch-free nature of adaptive schemes. Furthermore, we prove full sequence convergence without prior knowledge of local Hoölder constants nor of the order of Hölder continuity. Numerical experiments make comparisons with baseline methods on diverse tasks from machine learning covering both the locally and the globally Hölder setting.

Peter Richtárik (KAUST, Saudi Arabia)

The First Optimal Parallel SGD (in the Presence of Data, Compute and Communication Heterogeneity)

The design of efficient parallel/distributed optimization methods and tight analysis of their theoretical properties are important research endeavors. While minimax complexities are known for sequential optimization methods, the theory of parallel optimization methods is surprisingly much less explored, especially in the presence of data, compute and communication heterogeneity. In the first part of the talk [1], we establish the first optimal time complexities for parallel optimization methods (Rennala SGD and Malenia SGD) that have access to an unbiased stochastic gradient oracle with bounded variance, under the assumption that the workers compute stochastic gradients with different speeds, i.e., we assume compute heterogeneity. We prove lower bounds and develop optimal algorithms that attain them, both in the data homogeneous and heterogeneous regimes. In the second part of the talk [2], we establish the first optimal time complexities for parallel optimization methods (Shadowheart SGD) that have access to an unbiased stochastic gradient oracle with bounded variance, under the assumption that the workers compute stochastic gradients with different speeds, as before, but under the further assumption that the worker-to-server communication times are nonzero and heterogeneous. We prove lower bounds and develop optimal algorithms that attain them, in the data homogeneous regime only. Time permitting, I may briefly outline some related recent results [3, 4]. Our results have surprising consequences for the literature of asynchronous optimization methods: in contrast with prior attempts to tame compute heterogeneity via "complete/wild" compute and update asynchronicity, our methods alternate fast asynchronous computation of a minibatch of stochastic gradients with infrequent synchronous update steps.

[1] Alexander Tyurin and Peter Richtárik. Optimal time complexities of parallel stochastic optimization methods under a fixed computation model, Advances in Neural Information Processing Systems 36 (NeurIPS 2023).
[2] Alexander Tyurin, Marta Pozzi, Ivan Ilin, and Peter Richtárik. Shadowheart SGD: Distributed asynchronous SGD with optimal time complexity under arbitrary computation and communication heterogeneity, arXiv:2402.04785, 2024.
[3] Alexander Tyurin, Kaja Gruntkowska, and Peter Richtárik. Freya PAGE: First optimal time complexity for large-scale nonconvex finite-sum optimization with heterogeneous asynchronous computations, arXiv:2405.15545, 2024.
[4] Alexander Tyurin and Peter Richtárik. On the optimal time complexities in decentralized stochastic asynchronous optimization, arXiv:2405.16218, 2024.

Anton Rodomanov (CISPA Helmholtz Center for Information Security, Germany)

Universality of AdaGrad Stepsizes for Stochastic Optimization

We present adaptive gradient methods (both basic and accelerated) for solving convex composite optimization problems in which the main part is approximately smooth (a.k.a. (δ, L)-smooth) and can be accessed only via a (potentially biased) stochastic gradient oracle. This setting covers many interesting examples including Hölder smooth problems and various inexact computations of the stochastic gradient. Our methods use AdaGrad stepsizes and are adaptive in the sense that they do not require knowing any problem-dependent constants except an estimate of the diameter of the feasible set but nevertheless achieve the best possible convergence rates as if they knew the corresponding constants. We demonstrate that AdaGrad stepsizes work in a variety of situations by proving, in a unified manner, three types of new results. First, we establish efficiency guarantees for our methods in the classical setting where the oracle's variance is uniformly bounded. We then show that, under more refined assumptions on the variance, the same methods without any modifications enjoy implicit variance reduction properties allowing us to express their complexity estimates in terms of the variance only at the minimizer. Finally, we show how to incorporate explicit SVRG-type variance reduction into our methods and obtain even faster algorithms. In all three cases, we present both basic and accelerated algorithms achieving state-of-the-art complexity bounds. As a direct corollary of our results, we obtain universal stochastic gradient methods for Hölder smooth problems which can be used in all situations.

Kees Roos (Delft University of Technology, Netherlands)

Search directions for primal-dual interior-point methods

Primal-dual IPMs are known for their efficiency. There exist two wide classes of search directions for these methods. One is the class of methods that use 'kernel functions', that have their origin in the popular logarithmic barrier function, and the other class is based on the 'algebraically equivalent transformation' (shortly AET) approach. We discuss the ideas underlying both approaches and their relation.

Katya Scheinberg (Cornell University, USA)

Stochastic oracles and where to find them

Majority of modern optimization methods, from stochastic gradient descent to "zero-th order" methods use some kind of approximate first order information. We will overview different methods of obtaining this information including simple stochastic gradient via sampling, robust gradient estimation in adversarial settings, traditional and randomized finite difference methods and more. We will discuss what key properties of these inexact, stochastic first order oracles are useful for convergence analysis of optimization methods that use them.

Vladimir Shikhman (Chemnitz University of Technology, Germany)

Dynamic pricing under discrete choice demand

Recently, there is growing interest and need for dynamic pricing algorithms, especially, in the field of online marketplaces by offering smart pricing options for big online stores. We present an approach to adjust prices based on the observed online market data. The key idea is to characterize optimal prices as minimizers of a total expected revenue function, which turns out to be convex. We assume that consumers face information processing costs, hence, follow a discrete choice demand model, and suppliers are equipped with quantity adjustment costs. We prove the strong smoothness of the total expected revenue function by deriving the strong convexity modulus of its dual. Our gradient-based pricing schemes outbalance supply and demand at the optimal convergence rates. This suggests that the imperfect behavior of consumers and suppliers helps to stabilize the market.

In collaboration with David Müller and Yurii Nesterov.

Damien Scieur (Samsung - SAIL Montréal, Canada)

Adaptive Quasi-Newton and Anderson acceleration framework with explicit global (accelerated) convergence rates

Despite the impressive numerical performance of the quasi-Newton and Anderson/nonlinear acceleration methods, their global convergence rates have remained elusive for over 50 years. This study addresses this long-standing issue by introducing a framework that derives novel, adaptive quasi-Newton and nonlinear/Anderson acceleration schemes. Under mild assumptions, the proposed iterative methods exhibit explicit, non-asymptotic convergence rates that blend those of the gradient descent and Cubic Regularized Newton's methods. The proposed approach also includes an accelerated version for convex functions. Notably, these rates are achieved adaptively without prior knowledge of the function's parameters. The framework presented in this study is generic, and its special cases include algorithms such as Newton's method with random subspaces, finite differences, or lazy Hessian. Numerical experiments demonstrated the efficiency of the proposed framework, even compared to the l-BFGS algorithm with Wolfe line-search.

Vladimir Spokoiny (WIAS, Germany)

Uncertainty quantification under self-concordance

Statistical maximum-likelihood or M-estimation lead to optimization of a stochastic function. Statistical inference can be viewed as comparing the solutions of the original stochastic problem and its population counterpart. Under some conditions including self-concordance, we provide finite sample guarantees for loss and risk of such estimators.

Tamás Terlaky (Lehigh University, USA)

How we got to Quantum Interior Point Methods?

70 years ago, Ragnar Frish proposed a logarithmic potential method. 40 years ago, Karmarkar proposed a polynomial time potential reduction IPM, and a few years later Nesterov and Nemirovskii developed the general theory of polynomial barrier methods for convex optimization, including optimization over positive definite matrices. This work was crucial in launching the emerging wave of Quantum Interior Point Methods. The quest for faster algorithms, improving computational complexity goes on ...

Philippe Toint (Université de Namur, Belgium)

Some results in Objective-Function Free Optimization (OFFO)

The talk will review the motivation for Objective-Function Free Optimization (OFFO) and some of the recent results. Starting with two parametric classes of first-order methods including the well-know Adagrad algorithm , it will outline an application to multi-level strategies for neural network training, Second-order methods (possibly using high-order derivatives) will then be investigated. Evaluation complexity results will be given for both cases, in the deterministic and stochastic contexts.

Levent Tunçel (University of Waterloo, Canada)

Some recent developments in primal-dual algorithms in algorithmic optimization and their connections to Yurii Nesterov's contributions

I will summarize some of the recent research results I have obtained with various co-authors in algorithmic optimization. These include new results concerning primal-dual interior-point methods as well as new results concerning polynomial-time primal-dual approximation algorithms for some NP-hard problems. Throughout the talk, I will point out many connections to Yurii Nesterov's contributions.

The new results to be reported in this talk include joint work with N. Benedetto Proenca, J. Dahl, M. de Carli Silva, M. Karimi, V. Roshchina, C. Sato and L Vandenberghe.

Luis Nunes Vicente (Lehigh University, USA)

Multi-objective optimization revisited: Most-changing Pareto sub-fronts

In this talk, we first review basic concepts for multi-objective optimization and summarize the main features of deterministic and stochastic multi-gradient descent. We will describe the properties of the multi-gradient and the associated complexity rates. We will see how to compute Pareto fronts for bi-objective learning problems where accuracy and fairness are conflicting goals. In the second part of the talk, using Pareto sensitivity, we will introduce a new approach to identify where the objective functions change the most on a Pareto front. We first see how to compute the most-changing Pareto neighborhoods around a Pareto solution, where points are distributed along directions of maximum change. We then see how to compute regions or sub-fronts of the entire Pareto front where the most change occurs. In doing so, we hope to offer a new perspective on how to compute representative Pareto solutions, in ways that are more comprehensive than existing techniques such as the so-called Pareto knee solutions. Our techniques are still restricted to scalarized methods, in particular to the weighted-sum or epsilon-constrained methods, and require the computation or approximations of first- and second-order derivatives.

This is joint work with Tommaso Giovannelli, Suyun Liu, and Marcos Medeiros Raimundo.

Yinyu Ye (Stanford University, USA)

Computational Progress on Mathematical Optimization Solvers

We describe some recent advances in the development of general-purpose numerical algorithms for linear, semidefinite programming (LP/SDP), and AI Training optimization, including fast pre-solving and diagonal-preconditioning, LP cross-over, ADMM, First-order Hype-gradients, GPU Implementations. Most of these techniques have been implemented in emerging optimization solvers, and they increased the average solution speed by over 3x in the past three years on a set of benchmark problems. For certain problem types, the speedup is more than 100x, and problems that have taken days to solve or never been solved before are now solved in seconds or minutes to high accuracy.

Poster Presentations

Numerical Performance of Nonmonotone line search Methods on Global Optimization Problems

Presenter: Zohreh Aminifard (UCLouvain, Belgium)

Abstract: Standard line search methods typically require that the function values of iterates decrease monotonically. While this condition seems reasonable and even desirable for minimization algorithms, it has two significant drawbacks when applied to nonconvex problems. First, enforcing monotonicity can slow the progress of the iterates, hindering the optimization process. Second, monotone methods often cause iterates to become stuck at local minima, which is a critical issue when minimizing functions with multiple non-global local minima. To address these challenges, nonmonotone line search methods can be employed. Unlike their monotone counterparts, these methods do not require a decrease in function value between consecutive iterates. This flexibility allows the iterates to "jump over hills," enabling them to escape from nearby non-global local minima. Recently, Grapiglia and Sachs (2017) established convergence and complexity results for a wide class of nonmonotone line search methods, paving the way for the development of various new nonmonotone methods with worst-case complexity guarantees for finding approximate stationary points. In this work, we evaluate the numerical performance of these methods through extensive experiments on benchmark global optimization problems.

The inexact power augmented Lagrangian method for constrained nonconvex optimization

Presenter: Alexander Bodard (KU Leuven)

Abstract: We analyze an inexact augmented Lagrangian method with an unconventional augmenting term for solving a general class of nonconvex problems with nonlinear constraints. It is shown that this method achieves a better iteration complexity in terms of the constraint violation than its classical counterpart. Additionally, we provide a total complexity analysis when using an accelerated first-order method to solve the Hölder-smooth inner problems. This result suggests that the optimal power of the augmenting term depends on the desired tolerance for the suboptimality and constraint violation. Finally, numerical experiments validate the practical merits of our power augmented Lagrangian method on some problems of practical interest.

Worst-case analysis of first-order optimization methods involving linear operators

Presenter: Nizar Bousselmi (UCLouvain, Belgium)

Abstract: The Performance Estimation Problem (PEP) framework allows to automatically compute the worst-case performance of a given first-order optimization method on a given function class. We develop a PEP representation for linear operators. We consider convex optimization problems involving linear operators, such as those whose objective functions include compositions by a linear operator or featuring linear constraints. Our goal is to identify, thanks to PEP, the worst-case behavior of methods designed to solve such problems. We demonstrate the scope of our tool by computing several tight worst-case convergence rates, including the gradient method and the Chambolle-Pock method.

A derivative-free trust-region method based on finite differences for composite nonsmooth optimization

Presenter: Dânâ Davar (UCLouvain, Belgium)

Abstract: We present a derivative-free trust-region method based on finite differences for minimizing composite functions of the form $h(F(x))$, where h is a convex Lipschitz function, possibly nonsmooth, while F is a vector blackbox function with Lipschitz continuous Jacobian. The proposed method approximates the Jacobian of F by forward finite differences with a dynamic stepsize. It is shown that the method needs at most $\mathcal{O}(n\epsilon^{-2})$ function evaluations to find an ϵ-approximate stationary point. Numerical results are also reported, showing the relative efficiency of the new method with respect to state-of-the-art derivative-free solvers for composite nonsmooth optimization.

Exact Worst-case Convergence Rate for Primal Solutions of Dual Methods Using Performance Estimation Problem

Presenter: Zhicheng Deng (ShanghaiTech University and UCLouvain, Belgium)

Abstract: In this poster, we present a tight non-ergodic convergence rate analysis for the primal solutions of the dual methods including dual ascent, augmented Lagrangian using Performance Estimation Problem and duality theory. Compared to the existing literature which conducts the proof relying on the Lagrangian, our approach focuses on the dual aspect of the optimization problem, which leads to a conceptually simpler representation for the algorithm and reduces to the traditional function class and algorithms. We give a conjecture on the measure of non-ergodic rate for the primal objective gap of the dual ascent. Moreover, we show how to expand our methodology to a wider range of dual methods.

Progressive decoupling of linkage problems beyond elicitable monotonicity

Presenter: Brecht Evens (KU Leuven, Belgium)

Abstract: We study the progressive decoupling algorithm and Spingarn's method of partial inverses for solving the linkage problem, which involves finding a point in the graph of an operator with the primal variable in a closed subspace and the dual variable in its orthogonal complement. This problem is relevant to optimization and variational inequalities due to its connection with Lagrangian duality. We establish convergence of both methods without requiring monotonicity assumptions, by leveraging their connection with the preconditioned proximal point method. Additionally, we illustrate the applicability of our theory with several examples.

Nonsmooth trust-region methods via Bregman forward-backward envelope

Presenter: Mohammad Hamed (University of Antwerp, Belgium)

Abstract: Objective functions exhibiting relative smoothness—a broader concept than Lipschitz smoothness—are widely applicable in areas like signal and image processing, machine learning, and inverse problems. In this research, we develop a trust-region method, which is a smooth optimization technique, to obtain a critical point of the original function. This is achieved by creating a smooth approximation of the objective function using the Bregman forward-backward envelope (BFBE). We provide both the convergence analysis and numerical experiments, highlighting the theoretical and practical efficiency of the proposed method.

Block Majorization Minimization with Extrapolation and Application to β-NMF

Presenter: Nicolas Gillis (University of Mons, Belgium)

Abstract: In this poster, we present a new Block Majorization Minimization method with Extrapolation (BMMe) for solving a class of multi-convex optimization problems. The extrapolation parameters of BMMe are updated using a novel adaptive update rule. By showing that block majorization minimization can be reformulated as a block mirror descent method, with the Bregman divergence adaptively updated at each iteration, we establish subsequential convergence for BMMe. We use this method to design state-of-the-art algorithms to tackle nonnegative matrix factorization problems with the β-divergences (β-NMF) for $\beta \in [1, 2]$. These algorithms, which are multiplicative updates with extrapolation, benefit from our novel results that offer convergence guarantees. We also empirically illustrate the significant acceleration of BMMe for β-NMF through extensive experiments. The paper is available from `https://arxiv.org/abs/2401.06646`

Scaled Gradient Algorithm Meets Local Smoothness

Presenter: Susan Ghaderi (KU Leuven, Belgium)

Abstract: In this poster, we introduce a scaled gradient algorithm for smooth convex functions with a locally Lipschitz gradient. To the best of our knowledge, this is the first algorithm of its kind introduced for this class of problems. We studied the convergence and complexity analysis under suitable assumptions. Furthermore, we prove the linear convergence of our algorithm under the locally strong convexity of the function and Łojasiewicz inequalities. Comprehensive numerical experiments are conducted, showcasing the effectiveness and efficiency of the proposed method compared to standard gradient and adaptive gradient methods.

High-Order Proximal-Point Methods in the Nonconvex Setting

Presenter: Alireza Kabgani (University of Antwerp, Belgium)

Abstract: This poster explores the study of inexact high-order proximal-point methods in the realm of nonconvex optimization. Our focus is on demonstrating subsequential convergence by showing the monotonic decrease of Lyapunov function values, using the high-order Moreau envelope as our Lyapunov function. We highlight the effectiveness of these algorithms in addressing weakly convex optimization problems, supported by comprehensive theoretical analyses and numerical results, particularly in solving robust low-rank matrix recovery problems.

Numerical design of optimized first-order methods

Presenter: Yassine Kamri (UCLouvain, Belgium)

Abstract: Large-scale optimization is central to many engineering applications such as machine learning, signal processing, and control. First-order optimization methods are a popular choice for solving such problems due to their performance and low computational cost. However, in many applications, such as machine learning, their performance is highly dependent on the choice of step sizes. Tuning the step sizes of first-order methods is a notoriously hard nonconvex problem. In this talk, we present several approaches to optimizing the step sizes of first-order optimization methods, relying on recent developments in Performance Estimation Methods. We compare our results to existing optimized step sizes and provide new results, namely numerically optimized steps for cyclic coordinate descent and inexact gradient descent, leading to accelerated convergence rates.

Optimizing Recommender Systems: Balancing Accuracy and Diversity

Presenter: Elaheh Lotfian (Independent Researcher)

Abstract: This poster presents a multi-objective recommender system aimed at enhancing both accuracy and diversity in top-n recommendation lists. We introduce a customized hybrid AMOSA_NSGA-II (HAN) algorithm that generates a Pareto set of top-s lists and a methodology for selecting the optimal list for each user from this set. Our approach begins with the generation of a preliminary top-k list through item-based collaborative filtering. In the second stage, we address the bi-objective optimization problem using the HAN algorithm. The final stage involves creating an optimal personalized top-s list tailored to each individual user. To assess the effectiveness of our approach, we apply it to real-world datasets, conduct a comprehensive performance evaluation, and compare our results with existing methods in the literature.

Learning the subspace of variation for global optimization of functions with low effective dimension

Presenter: Estelle Massart (UCLouvain, Belgium)

Abstract: We propose an algorithmic framework that employs active subspace techniques for scalable global optimization of functions with low effective dimension (also referred to as low-rank functions). These functions arise, e.g., in hyperparameter tuning for deep neural networks and in large-scale engineering and physical simulations models. This proposal replaces the original high-dimensional problem by one or several lower-dimensional reduced subproblem(s), capturing the main directions of variation of the objective which are estimated here as the principal components of a collection of sampled gradients. We quantify the sampling complexity of estimating the subspace of variation of the objective in terms of its effective dimension and hence, bound the probability that the reduced problem will provide a solution to the original problem. To account for the practical case when the effective dimension is not known a priori, our framework adaptively solves a succession of reduced problems, increasing the number of sampled gradients until the estimated subspace of variation remains unchanged. We prove global convergence under mild assumptions on the objective, the sampling distribution and the subproblem solver, and illustrate numerically the benefits of our proposed algorithms over those using random embeddings. This is joint work with C. Cartis (University of Oxford), X. Liang (University of Manchester) and A. Otemissov (Nazarbayev University).

Linearly convergent subgradient method for paraconvex optimization

Presenter: Morteza Rahimi (University of Antwerp, Belgium)

Abstract: The primary aim of this work is to analyze the convergence of subgradient methods applied to paraconvex functions that adhere to a specific error bound property. We demonstrate that these subgradient methods achieve linear convergence in the context of paraconvex functions. Notably, the parameters related to paraconvexity and error bounds are only relevant for determining an initial valid point. Consequently, the linear convergence rate is independent of these parameters. Preliminary numerical results support our theoretical findings. This is a joint work with Prof. Dr. Masoud Ahookhosh.

Forward backward envelopes under the lens of generalized convexity: Unifying framework and algorithms

Presenter: Konstantinos Oikonomidis (KU Leuven, Belgium)

Abstract: We present an abstract forward-backward splitting algorithm that can be interpreted in terms of a generalized convex-concave procedure. The algorithm unifies many popular existing methods such as the classical convex-concave procedure, Bregman and classical forward-backward splitting, proximal point algorithms with phi-divergences, natural gradient descent and motivates entirely new algorithms such as an anisotropic generalization of forward-backward splitting. We analyze the algorithms' sublinear and linear convergence under an abstract proximal PL-inequality. In the spirit of the classical convex-concave procedure we also study certain equivalent reformulations of the optimization problem involving generalized conjugate functions obtained via a double-min duality. This viewpoint leads to a natural generalization of the Moreau and forward-backward envelope (called Phi-conjugates in the context of generalized convexity) that serve as real-valued continuous surrogates preserving (local) minima and stationary points. Leveraging the generalized implicit function theorem we study the 1st and 2nd order-differentiability properties of these envelopes. Ultimately this is used to obtain globalized (quasi-)Newton accelerated versions of the aforementioned algorithms.

Multi-Objective Optimization in Machine Learning - Exploring Conflicts in Learning

Presenter: Marcos M. Raimundo (Unicamp, Brazil)

Abstract: Creating machine learning models is an inherently conflicting process. From the bias vs. variance trade-off, we observe that reducing learning error often increases model complexity. Thus, error and complexity can be considered a pair of conflicting objectives among many others in machine learning: false positive rate vs. false negative rate, accuracy vs. model fairness, accuracy vs. model interpretability. In this context, it is appropriate to treat such formulations as multi-objective optimization problems and use methodologies that explore multiple efficient learning models, each characterized by a distinct trade-off between conflicting objectives. These multiple models found through multi-objective optimization can create options for the decision maker's preferences, inducing models with different data views and generating diversity among the models. With this in mind, we can think of a multi-objective learning conceptual framework designed by three steps: (1) Multi-objective modeling of each learning problem, express the conflicting objectives; (2) Search for efficient and well-distributed solutions along the Pareto frontier with optimization solvers; (3) Post-selection among efficient models using the decision maker's experience or filtering and aggregation for ensemble synthesis. This framework can contribute to various machine learning problems, including (i) multi-class classification, (ii) imbalanced classification, (iii) multi-label classification, (iv) multi-task learning, (v) learning with multiple feature sets, (vi) reduction of discriminatory bias, and (vii) improving explainability and trustworthiness.

Improved convergence rates for the Difference-of-Convex algorithm

Presenter: Teodor Rotaru (KU Leuven, Belgium)

Abstract: We consider a difference-of-convex formulation where one of the terms is allowed to be hypoconvex (or weakly convex). We first examine the precise behavior of a single iteration of the Difference-of-Convex algorithm (DCA), giving a tight characterization of the objective function decrease. This requires distinguishing between eight distinct parameter regimes. Our proofs are inspired by the performance estimation framework, but are much simplified compared to similar previous work. We then derive sublinear DCA convergence rates towards critical points, distinguishing between cases where at least one of the functions is smooth and where both functions are nonsmooth. We conjecture the tightness of these rates for four parameter regimes, based on strong numerical evidence obtained via performance estimation, as well as the leading constant in the asymptotic sublinear rate for two more regimes.

Analysis of second-order methods, via non-convex Performance Estimation framework and tight representation of second-order function classes

Presenter: Anne Rubbens (UCLouvain, Belgium)

Abstract: The "Performance Estimation Problem framework" computes numerically the exact worst-case performance of a given optimization method on a given function class. It allowed to determine the convergence rate of numerous first-order methods and also to design optimal methods. When Performance Estimation Problems can be formulated in a convex way, they can be efficiently solved, but these formulations are restricted the analysis to certain types of methods, in particular to many first-order methods and function classes. In this work, we extend the framework to analyze second-order methods, by formulating such analysis as a non convex problem and building on some recent improvements in solving non-convex Performance Estimation Problem. To obtain exact bounds on the performance of various methods, we also provide an exact description of various second-order classes of univariate functions, e.g. function with Lipschitz continuous Hessian or self-concordant functions.

Deep Asymmetric Nonnegative Matrix Factorization for Graph Clustering

Presenter: Amjad Seyedi (University of Mons, Belgium)

Abstract: This research investigates the challenges of directed graph clustering through the lens of nonnegative matrix factorization (NMF). We propose Deep Asymmetric NMF (DAsNMF), a novel method that leverages the strengths of both asymmetric NMF and deep learning for superior performance and interpretability. Our key innovation lies in formulating a unified optimization problem that integrates: 1) first- and second-order asymmetric similarity measures to enhance local structure preservation, 2) a deep factorization scheme for extracting hierarchical representations and enabling graph summarization, and 3) a graph regularization term incorporating global topological information. This joint optimization not only improves clustering accuracy but also provides insights into the multi-scale structure of directed graphs. We develop an efficient alternating minimization algorithm to solve the optimization problem and demonstrate its effectiveness through comprehensive experiments on real-world datasets. Our work bridges the gap between deep learning and graph clustering, offering valuable tools for interpretable network analysis.

Inexact Levenberg-Marquardt method for Holder metrically subregular mappings

Presenter: Bas Symoens (University of Antwerp, Belgium)

Abstract: Many scientific and industrial applications involve solving systems of nonlinear equations $F(x) = 0$. Attempting to solve this problem leads to one of many iterative schemes, such as Levenberg-Marquardt: a combination of gradient descent and Gauss-Newton. Traditional Levenberg-Marquardt methods assume a nonsingular gradient for F, leading to a local error bound condition (the distance of the current iterate x_k to the solution set is smaller than some positive constant times $\|F(x_k)\|$). However, this assumption might be restrictive for many applications. To combat this, Hölder metric subregularity was introduced, a weaker condition than the local error bound, applicable to a broader range of functions. Our goal is to combine results for this generalisation, together with a more general (adaptive) choice of the regularisation parameter of Levenberg-Marquardt. Solving the linear systems in each iteration of Levenberg-Marquardt can be challenging. We address this by considering inexact iterative solvers and propose accuracy conditions to ensure local convergence. The choice of the proper iterative solver and ways to adapt them to the specific case of a Levenberg-Marquardt system is discussed and illustrated with examples.

Enhancing finite-difference based derivative-free optimization methods with Sobolev trained surrogate models

Presenter: Timothé Taminiau (UCLouvain, Belgium)

Abstract: Derivative-free optimization (DFO) aims to minimize a function using only zero-order information. The complexity of DFO methods is driven by the number of function evaluations, which are computationally expensive for many applications of interest. We investigate the possibility of reducing the number of function evaluations required by the algorithm to achieve a given accuracy via machine learning based models of the objective function. In particular, we consider shallow neural network approximation and kernel interpolation models. We show that the extrapolation capacity is improved when these models fit both previously computed function values and gradient approximations, known as Sobolev training. This is joint work with Estelle Massart and Geovani Grapiglia (UCLouvain)

A principled approach to automatically recommend optimization methods

Presenter: Sofiane Tanji (UCLouvain, Belgium)

Abstract: As the literature on optimization algorithms has been ever-growing, a wide range of methods are now available. Usually, published algorithms are shown to be applicable to given templates of optimization problems, and are often accompanied by proofs characterizing their convergence rates. Many such templates exist and are not always clearly identifiable. Hence, given an optimization problem, it is often hard to identify which templates it can match, especially if one also wants to consider equivalent reformulations of the problem. This makes the task of choosing an optimization method tedious. We introduce a framework to automatically (1) check whether an optimization method is applicable to a user-provided optimization problem and (2) rank the best applicable methods according to their worst-case theoretical guarantees. The proposed framework, implemented as a Python package, achieves this via an extensive library of optimization methods with their associated known convergence rates.

Continuous Dynamics for Discrete Optimization via Kuramoto-type Oscillators with Exact Penalty and Generalized Coupling

Presenter: Guanchun Tong (KU Leuven, Belgium)

Abstract: We discuss a dynamical systems perspective on discrete optimization. Departing from the fact that many combinatorial optimization problems can be reformulated as finding low energy spin configurations in corresponding Ising models, we derive a penalized rank-two relaxation of the Ising formulation. It turns out that the associated gradient flow dynamics exactly correspond to a type of hardware solvers termed oscillator-based Ising machines. We also analyze the advantage of adding penalties by leveraging random rounding techniques. Therefore, our work contributes to a rigorous understanding of the oscillator-based Ising machines by drawing connections to the penalty method in constrained optimization and providing a rationale for the introduction of sub-harmonic injection locking mechanism (as exact penalty). Furthermore, we characterize a class of generalized coupling functions between oscillators, which ensures convergence to discrete solutions. This class of coupling functions therefore avoids explicit penalty terms or rounding schemes, which are prevalent in other formulations.

A Riemannian rank-adaptive method for tensor completion in the tensor-train format

Presenter: Charlotte Vermeylen (KU Leuven, Belgium)

Abstract: Riemannian rank-adaptive methods (RRAMs) are state-of-the-art methods that aim to optimize a continuously differentiable function on a low-rank variety, a problem appearing notably in low-rank matrix or tensor completion. The RRAMs that are developed for the set of bounded rank matrices iteratively optimize over smooth fixed-rank manifolds, starting from a low initial rank. They increase the rank by performing a line search along a descent direction selected in the tangent cone to the variety. This direction can be the projection of the negative gradient onto the tangent cone but does not need to; for instance, the RRAM developed by Gao and Absil uses the projection of the negative gradient onto the part of the tangent cone that is normal to the tangent space. Additionally, they decrease the rank based on a truncated SVD when the algorithm converges to an element of a lower-rank set. This is possible because the manifold is not closed. We aim to generalize these RRAMs to the tensor-train (TT) format. In this format, only the RRAM by Steinlechner is known to us from the literature. This method is developed for high-dimensional tensor completion and has a random rank update mechanism in the sense that each TT-rank is increased subsequently by one by adding a small random term in the TTD of the current best approximation such that the cost function does not increase much. No rank reduction step is included which makes the algorithm prone to overfitting. We improve this RRAM by including a method to increase the TT-rank based on an approximate projection of the negative gradient onto the tangent cone to the variety. The tangent cone is the set of all tangent vectors to the variety. The approximate projection ensures that the search direction is sufficiently gradient related. Furthermore, a method is proposed to determine how much the rank should be increased for the low-rank tensor completion problem (LRTCP). Lastly, a method to decrease the rank is proposed, which is necessary when the iterate comes close to a lower-rank set. This is possible because, as for the manifold of fixed-rank matrices, the manifold of fixed-rank TTDs is not closed.

A quadratically convergent proximal algorithm for nonnegative tensor decomposition

Presenter: Nico Vervliet (KU Leuven, Belgium)

Abstract: The decomposition of tensors into simple rank-1 terms is key in a variety of applications in signal processing, data analysis and machine learning. While this canonical polyadic decomposition (CPD) is unique under mild conditions, including prior knowledge such as nonnegativity can facilitate interpretation of the components. Inspired by the effectiveness and efficiency of Gauss–Newton (GN) for unconstrained CPD, we derive a proximal, semismooth GN type algorithm for nonnegative tensor factorization. Global convergence to local minima is achieved via backtracking on the forward-backward envelope function. If the algorithm converges to a global optimum, we show that Q-quadratic rates are obtained in the exact case. Such fast rates are verified experimentally, and we illustrate that using the GN step significantly reduces number of (expensive) gradient computations compared to proximal gradient descent.

Benign nonconvexity and global synchronization with nonlinear interactions

Presenter: Hongjin Wu (ETH Zürich, Switzerland)

Abstract: Nonlinear dynamics and nonconvex optimization are closely related. We consider the dynamics of n points on a unit circle where each connected pair of nodes prefers to be positioned at the same angle. These dynamics are equivalent to seeking the minima of the energy $E(\theta_1, \ldots, \theta_n) = -\sum_{i,j=1}^{n} a_{ij} \cos(\theta_i - \theta_j)$, which can be viewed as a rank-two Burer-Monteiro factorization of $\mathbb{Z}_2$ synchronization [Bandeira et. al., 2016, Burer and Monteiro, 2003]. All second-order critical points of this energy landscape on certain graphs are global minima, representing a synchronization state - this is known as benign nonconvexity and the global synchronization phenomenon. But does the coupling function have to be the cosine function? In this work, we show that in a certain regime of nonlinearity of $\phi(x)$, the energy $E(\theta_1, \ldots, \theta_n) = -\sum_{i,j=1}^{n} a_{ij} \phi(\cos(\theta_i - \theta_j))$ remains benign, with global minima representing synchronization states. As a specific example, we consider the coupling function $\phi_\beta(x) = \frac{1}{\beta} e^{\beta x}$, which describes the dynamics of idealized transformers [Geshkovski, 2023]. Our results imply that global synchronization holds in the regime $\beta \lesssim n^{-1/3}$.

The exact convergence rate of the subgradient method with Polyak step size

Presenter: Moslem Zamani (UCLouvain, Belgium)

Abstract: This poster investigates the subgradient method employing the Polyak step size for minimizing nonsmooth convex functions with bounded subgradients. Our analysis demonstrates that the traditional subgradient method with Polyak step size achieves a convergence rate of $\mathcal{O}\left(\frac{1}{\sqrt[4]{N}}\right)$ in terms of the final iterate, and we provide an example illustrating the exactness of this rate. We also introduce an adaptive Polyak step size, which its convergence rate matches the theoretical lower bound. Furthermore, we propose an adaptive Polyak method with a momentum term, featuring step sizes that are independent of the iteration count. We prove that the convergence rate of the method matches the lower bound, too. Additionally, we explore the alternating projection method and derive an exact convergence rate.

Automated tight Lyapunov analysis for first-order methods

Presenter: Manu Upadhyaya (Lund University, Sweden)

Abstract: We present a methodology for establishing the existence of quadratic Lyapunov inequalities for a wide range of first-order methods used to solve convex optimization problems. In particular, we consider (i) classes of optimization problems of finite-sum form with (possibly strongly) convex and possibly smooth functional components, (ii) first-order methods that can be written as a linear system on state-space form in feedback interconnection with the subdifferentials of the functional components of the objective function, and (iii) quadratic Lyapunov inequalities that can be used to draw convergence conclusions. We present a necessary and sufficient condition for the existence of a quadratic Lyapunov inequality within a predefined class of Lyapunov inequalities, which amounts to solving a small-sized semidefinite program. We showcase our methodology on several first-order methods that fit the framework. Most notably, our methodology allows us to significantly extend the region of parameter choices that allow for duality gap convergence in the Chambolle–Pock method when the linear operator is the identity mapping.

Acceleration Exists! Optimization Problems When Oracle Can Only Compare Objective Function Values

Presenter: Aleksandr Lobanov (Moscow Institute of Physics and Technology, Russia)

Abstract: Frequently, the burgeoning field of black-box optimization encounters challenges due to a limited understanding of the mechanisms of the objective function. To address such problems, in this work we focus on the deterministic concept of Order Oracle, which only utilizes order access between function values (possibly with some bounded noise), but without assuming access to their values. As theoretical results, we propose a new approach to create non-accelerated optimization algorithms (obtained by integrating Order Oracle into existing optimization "tools") in non-convex, convex, and strongly convex settings that are as good as both SOTA coordinate algorithms with first-order oracle and SOTA algorithms with Order Oracle up to logarithm factor. Moreover, using the proposed approach, we provide the first accelerated optimization algorithm using the Order Oracle. And also, using an already different approach we provide the asymptotic convergence of the first algorithm with the stochastic Order Oracle concept. Finally, our theoretical results demonstrate effectiveness of proposed algorithms through numerical experiments. Link: `https://arxiv.org/pdf/2402.09014`.

Distributed Feedback Optimization of Linear Multi-agent Systems

Presenter: Amir Mehrnoosh (UCLouvain, Belgium)

Abstract: Feedback optimization has evolved as an important control approach for optimizing dynamical systems, particularly in terms of steady-state performance objectives. Traditional feedback optimization methods are primarily reliant on centralized systems and controller structures, which suffer from scalability and privacy issues for large-scale applications. This work investigates a distributed feedback optimization framework in which each agent updates its local control state by averaging inputs from neighboring agents and performing a local negative gradient step. Assuming convexity and smoothness of the cost, we show that this control technique converges to a fixed point. Further, under the weaker assumption of restricted strong convexity of the cost function, we prove that the algorithm achieves linear convergence to a vicinity of the optimal point, with the neighborhood size contingent on the chosen stepsize. Simulation results support the theoretical findings.

List of participants

Pierre-Antoine Absil
Masoud Ahookhosh
Zohreh Aminifard
Francis Bach
Michel Baes
Vincent Blondel
Alexander Bodard
Nicolas Boumal
Nizar Bousselmi
Coralia Cartis
Dânâ Davar
Etienne de Klerk
Lieven De Lathauwer
Séverine De Visscher
Pierre Dehez
Zhicheng Deng
Olivier Devolder
Nikita Doikov
Radu Dragomir
Pavel Dvurechensky
Brecht Evens
Mohammad Hamed Firouzehtarash
Mihai Florea
Robert Freund
Yves Genin
Catherine Germain
Michel Gevers
Nicolas Gillis
François Glineur
Geovani Grapiglia
Susan Ghaderi
Nancy Guillaume
Osman Guler
Yvan Hachez
Niao He
Julien Hendrickx
Tibor Illés
Anatoli Iouditski
Miloica Jaćimović
Raphaël Jungers
Alireza Kabgani
Yassine Kamri
Guanghui (George) Lan
Goran Lesaja
Aleksandr Lobanov
Elaheh Lotfian
Estelle Massart
Simon Mataigne
Amir Mehrnoosh
Boris Mordukhovich
Ion Necoara
Arkadi Nemirovski
Yurii Nesterov
Konstantinos Oikonomidi
Panos Patrinos
Morteza Rahimi
Marcos Medeiros Raimundo
Peter Richtárik
Stefan Richter
Anton Rodomanov
Kees Roos
Teodor Rotaru
Anne Rubbens
Annick Sartenaer
Damien Scieur
Seyed Amjad Seyedi
Katya Sheinberg
Vladimir Shikhman
Vladimir Spokoiny
Bas Symoens
Timothé Taminiau
Firouzehtarash Sofiane Tanji
Tamás Terlaky
Philippe Toint
Guanchun Tong
Levent Tuncel
Manu Upadhyaya
Paul Van Dooren
Charlotte Vermeylen
Nico Vervliet
Luis Nunes Vicente
Laurence Wolsey
Hongjin Wu
Yinyu Ye
Moslem Zamani